

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 27, 2005, 17:31:33 ; Search time 69.2093 Seconds
(without alignments)
111.765 Million cell updates/sec

Title: US-10-612-090-14
Perfect score: 104
Sequence: 1 SSVLDVGYSPNREPLTGNS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	104	100.0	20	8	ADJ66741	Adj66741 Human Muc
2	104	100.0	95	8	ADP81172	ADP81172 Protein o
3	104	100.0	108	8	ADJ66729	Adj66729 Human Muc
4	104	100.0	108	8	ADJ66740	Adj66740 Human Muc
5	104	100.0	109	8	ADP81173	Adp81173 Protein o
6	104	100.0	174	7	ADM30756	Adm30756 Human can
7	104	100.0	284	6	ABU54859	Abu54859 Human CA1
8	104	100.0	438	5	AAE12623	AAe12623 Human gen
9	104	100.0	438	5	ABP30973	Abp30973 Partial p
10	104	100.0	438	7	ADA08636	Ada08636 Human O77
11	104	100.0	438	7	ADF08979	Adf08979 Secreted
12	104	100.0	526	6	ABU54709	Abu54709 Human CA1
13	104	100.0	526	4	AAE12634	AAe12634 Human gen
14	104	100.0	748	8	ADS94302	Ads94302 CA 125/O7
15	104	100.0	809	8	ADS94303	Ads94303 CA 125/O7
16	104	100.0	1148	4	ABBS0283	Abbs0283 Human pro
17	104	100.0	1148	4	ABBS0283	Abbs0283 HOST-1 ov
18	104	100.0	1148	5	ABG36380	Abg36380 Human ova
19	104	100.0	1148	5	ABP30964	Abp30964 Truncated
20	104	100.0	1148	5	ABP30969	Abp30969 Clone FLJ
21	104	100.0	1148	6	ABU54692	Abu54692 Human CA1
22	104	100.0	1148	6	ABU00145	Abu00145 Human nov
23	104	100.0	1148	7	ADA08632	Ada08632 Human O77
24	104	100.0	1148	7	ADA08611	Ada08611 Human O77
25	104	100.0	1148	7	ADF08954	Adf08954 Secreted

26	104	100.0	1148	7	ADF08975	Adf08975 Secreted
27	104	100.0	1148	7	ADG46241	Adg46241 Human ova
28	104	100.0	1148	7	ADM30754	Adm30754 Human can
29	104	100.0	1148	8	ADH41326	Adh41326 Human ova
30	104	100.0	1148	8	ADQ29705	Adq29705 Human col
31	104	100.0	1156	5	ABP30965	Abp30965 Protein w
32	104	100.0	1156	7	ADA08612	Ada08612 Human O77
33	104	100.0	1156	7	ADF08955	Adf08955 Secreted
34	104	100.0	1156	7	ADG46242	Adg46242 Human ova
35	104	100.0	1107	8	ADP10190	Adp10190 Human pro
36	104	100.0	1889	6	AAE34700	AAe34700 Protein e
37	104	100.0	1890	5	ABG36381	Abg36381 Human ova
38	104	100.0	1890	8	ADM12352	Adm12352 Human CA1
39	104	100.0	2234	8	ADJ66747	Adj66747 Human Muc
40	104	100.0	2248	6	ABU54693	Abu54693 Human CA1
41	104	100.0	11721	6	ABU54721	Abu54721 Human CA1
42	104	100.0	22152	8	ADR72871	Adr72871 Human ova
43	104	100.0	22157	8	ADP84155	Adp84155 Human CA1
44	101	97.1	1890	8	ADO38574	Ado38574 Ovarian c
45	100	96.2	318	5	ABP31025	Abp31025 Amino aci

ALIGNMENTS

RESULT 1	
ADJ66741	
ID	ADJ66741 standard; peptide; 20 AA.
AC	ADJ66741;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Human Muc1/Muc16 protein-related peptide SegID14.
XX	
KW	monoclonal antibody; epitope; non-shed extracellular portion;
KW	shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
KW	breast cancer; ovarian cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO2004005470-A2.
PD	
XX	
PD	15-JAN-2004.
XX	
PF	03-JUL-2003; 2003WO-US020907.
XX	
PR	03-JUL-2002; 2002US-0393094P.
XX	
PA	(IMMU-) IMMUNOGEN INC.
XX	
PI	Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
XX	
PI	Water CA;
XX	
DR	WPI; 2004-091350/09.
XX	
PT	New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
XX	
PT	treating or monitoring malignancies, such as breast or ovarian cancer.
XX	
XX	Claim 14; SEQ ID NO 14; 113pp; English.
XX	
CC	This invention relates to a novel isolated monoclonal antibody that
CC	specifically binds to an epitope of a non-shed extracellular portion of a
CC	shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC	useful for the development of compounds with a cytostatic or for gene
CC	therapy. The composition and methods are useful in diagnosing, treating
CC	or monitoring malignancies, such as breast or ovarian cancer. The present
CC	sequence is that of a peptide which was used for raising antibodies to
CC	the extracellular, non-shed region of Muc16 and which is claimed in the
XX	specification.
XX	
SQ	Sequence 20 AA;

Query Match 100.0%; Score 104; DB 8; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2,7e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSVLVGYSPPNRNEPLTGS 20
 |||||
 DB 1 SSVLVGYSPPNRNEPLTGS 20

RESULT 2
 ADP81172

ID ADP81172 standard; protein; 95 AA.

AC ADP81172;

DT 09-SEP-2004 (first entry)

DE Protein of human ovarian specific gene, SEQ ID No 206.

KM normal; neoplastic; ovarian; ovarian specific nucleic acid; OSNA;
 KM metastatic; cancer; vaccine; cytostatic; human.

OS Homo sapiens.

PN MO2004053079-A2.

PD 24-JUN-2004.

PF 08-DEC-2003; 2003MO-US038855.

PR 06-DEC-2002; 2002US-0431301P.

PR 30-JUN-2003; 2003US-0484584P.

PR 07-NOV-2003; 2003US-0518607P.

PA (D1AD-) DIADEXUS INC.

PI Macina RA, Turner LR, Sun Y, Liu S, Chen H;

DR WPI; 2004-468850/44.

DR N-PSDB; ADP81027.

PT New ovarian specific nucleic acid molecules and polypeptides useful for
 PT diagnosing, preventing or treating ovarian cancer, for producing
 PT transgenic animals or cells, or for research purposes.

PS Claim 12; SEQ ID NO 206; 754pp; English.

CC The invention relates to novel isolated nucleic acid molecules and
 CC polypeptides present in normal and neoplastic ovarian cells. These
 CC comprise a nucleic acid sequence encoding any of the 167 amino acid
 CC sequences (e.g. 438, 237 or 233 amino acids) fully defined in the
 CC specification (SEQ. ID NOS: ADP81095 to ADP81261) and comprises any of
 CC the 128 nucleotide sequences (e.g. 4798, 1494 or 1691 bp) fully defined
 CC in the specification (SEQ. ID NOS: ADP80967 to ADP81094). The invention
 CC further comprises: a method for determining the presence of a ovarian
 CC specific nucleic acid (OSNA) in a sample; a vector comprising the above
 CC nucleic acid molecule; a host cell comprising the vector; a method for
 CC producing a polypeptide encoded by the above nucleic acid molecule; a
 CC polypeptide encoded by the nucleic acid molecule cited above; an antibody
 CC or its fragment that specifically binds to the above polypeptide; a
 CC method for determining the presence of an ovarian specific protein in a
 CC sample; a method for diagnosing or monitoring the presence and metastases
 CC of ovarian cancer in a patient; a kit for detecting a risk of cancer or
 CC presence of cancer in a patient, the kit comprising a means for
 CC determining the presence of the above nucleic acid molecule or
 CC polypeptide; a method of treating a patient with ovarian cancer; and a
 CC vaccine comprising the above polypeptide or nucleic acid encoding the
 CC polypeptide. The isolated nucleic acid molecules and polypeptides have
 CC cytostatic activity. The isolated polypeptides may be used to create a
 CC vaccine. The isolated nucleic acid molecules and polypeptides can be used
 CC for diagnosing or monitoring the presence and metastases of ovarian
 CC cancer and treating ovarian cancer. This sequence represents the protein

CC of an ovarian specific gene of the invention.
 XX
 XX Sequence 95 AA;

Query Match 100.0%; Score 104; DB 8; Length 95;
 Best Local Similarity 100.0%; Pred. No. 1,6e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSVLVGYSPPNRNEPLTGS 20
 |||||
 DB 17 SSVLVGYSPPNRNEPLTGS 36

RESULT 3
 ADJ6729

ID ADJ6729 standard; protein; 108 AA.

AC ADJ6729;

DT 06-MAY-2004 (first entry)

DE Human Muc16 epitope amino acid sequence.

KM monoclonal antibody; epitope; non-shed extracellular portion;

KM shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;

KM breast cancer; ovarian cancer.

OS Homo sapiens.

PN MO2004005470-A2.

PD 15-JAN-2004.

PF 03-JUL-2003; 2003MO-US020907.

PR 03-JUL-2002; 2002US-0393094P.

PA (IMMU-) IMMUNOGEN INC.

PI Payne G, Chittenden T, Goldmakher V, Chun P, Snelder-Mulready K,

DR WPI; 2004-091350/09.

PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
 PT treating or monitoring malignancies, such as breast or ovarian cancer.

PS Claim 13; SEQ ID NO 2; 113pp; English.

CC This invention relates to a novel isolated monoclonal antibody that
 CC specifically binds to an epitope of a non-shed extracellular portion of a
 CC shed antigen or of human Muc1 or Muc16 protein. The invention may be
 CC useful for the development of compounds with a cytostatic or for gene
 CC therapy. The composition and methods are useful in diagnosing, treating
 CC or monitoring malignancies, such as breast or ovarian cancer. The present
 CC sequence is that of a Muc epitope of the invention.

XX Sequence 108 AA;

Query Match 100.0%; Score 104; DB 8; Length 108;
 Best Local Similarity 100.0%; Pred. No. 1,7e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSVLVGYSPPNRNEPLTGS 20
 |||||
 DB 86 SSVLVGYSPPNRNEPLTGS 105

RESULT 4
 ADJ6740

ID ADJ6740 standard; protein; 108 AA.

AC ADJ6740;

XX 06-MAY-2004 (first entry)
XX
XX
DE Human Nucle6 GST fusion protein amino acid sequence SeqID13.
XX
XX monoclonal antibody; epitope; non-shed extracellular portion;
KW shed antigen; human; Nucle6; cytosolic; gene therapy; malignancy;
KW breast cancer; ovarian cancer.
XX
XX Homo sapiens.
XX
XX WO2004005470-A2.
XX
XX 15-JAN-2004.
XX
XX 03-JUL-2003; 2003WO-US020907.
XX
XX 03-JUL-2002; 2002US-0393094P.
XX
XX (IMMU-) IMMUNOGEN INC.
XX
XX Payne G, Chittenden T, Goldmacher V, Chun P, Sneider-Mulready K;
PI Vater CA;
XX
XX WPI; 2004-091350/09.
XX
XX New antibody to non-shed human Nucle1 or Nucle6, useful for diagnosing,
PT treating or monitoring malignancies, such as breast or ovarian cancer.
XX
XX Disclosure; SEQ ID NO 13; 113pp; English.
XX
XX This invention relates to a novel isolated monoclonal antibody that
CC specifically binds to an epitope of a non-shed extracellular portion of a
CC shed antigen or of human Nucle1 or Nucle6 protein. The invention may be
CC useful for the development of compounds with a cytostatic or for gene
CC therapy. The composition and methods are useful in diagnosing, treating
CC or monitoring malignancies, such as breast or ovarian cancer. The present
CC sequence is that of the human Nucle6 GST fusion protein which was used in
CC the exemplification of the invention.
XX
XX
XX Sequence 108 AA;
SQ
Query Match 100.0%; Score 104; DB 8; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSVLVGYSPPNNEPLTGS 20
DB 86 SSVLVGYSPPNNEPLTGS 105
RESULT 5
ADP81173
ID ADP81173 standard; protein; 109 AA.
XX
XX ADP81173;
XX
XX 09-SEP-2004 (first entry)
XX
XX Protein of human ovarian specific gene, SEQ ID NO 207.
XX
XX normal; neoplastic; ovarian; ovarian specific nucleic acid; OSNA;
KW metastatic; cancer; vaccine; cytosolic; human.
XX
XX Homo sapiens.
XX
XX WO2004053079-A2.
XX
XX 24-JUN-2004.
XX
XX 08-DEC-2003; 2003WO-US038855.
XX
XX 06-DEC-2002; 2002US-0431301P.
XX
XX

PR 06-DEC-2002; 2002US-0431321P.
PR 30-JUN-2003; 2003US-0484584P.
PR 07-NOV-2003; 2003US-0518607P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Macina RA, Turner LR, Sun Y, Liu S, Chen H;
PI N-PSDB; ADP81027.
XX
XX WPI; 2004-46850/44.
XX
XX N-PSDB; ADP81027.
XX
XX New ovarian specific nucleic acid molecules and polypeptides useful for
PT diagnosing, preventing or treating ovarian cancer, for producing
PT transgenic animals or cells, or for research purposes.
XX
XX
XX Claim 12; SEQ ID NO 207; 754pp; English.
XX
XX
XX The invention relates to novel isolated nucleic acid molecules and
CC polypeptides present in normal and neoplastic ovarian cells. These
CC comprise a nucleic acid sequence encoding any of the 167 amino acid
CC sequences (e.g. 438, 237 or 233 amino acids) fully defined in the
CC specification (SEQ. ID NOS: ADP81095 to ADP81261) and comprises any of
CC the 128 nucleotide sequences (e.g. 4798, 1494 or 1691 bp) fully defined
CC in the specification (SEQ. ID NOS: ADP80967 to ADP81094). The invention
CC further comprises: a method for determining the presence of an ovarian
CC specific nucleic acid (OSNA) in a sample; a vector comprising the above
CC nucleic acid molecule; a host cell comprising the vector; a method for
CC producing a polypeptide encoded by the above nucleic acid molecule; a
CC polypeptide encoded by the nucleic acid molecule cited above; an antibody
CC or its fragment that specifically binds to the above polypeptide; a
CC method for determining the presence of an ovarian specific protein in a
CC sample; a method for diagnosing or monitoring the presence and metastases
CC of ovarian cancer in a patient; a kit for detecting a risk of cancer or
CC presence of cancer in a patient; the kit comprising a means for
CC determining the presence of the above nucleic acid molecule or
CC polypeptide; a method of treating a patient with ovarian cancer; and a
CC vaccine comprising the above polypeptide or nucleic acid encoding the
CC polypeptide. The isolated nucleic acid molecules and polypeptides have
CC cytostatic activity. The isolated polypeptides may be used to create a
CC vaccine. The isolated nucleic acid molecules and polypeptides can be used
CC for diagnosing or monitoring the presence and metastases of ovarian
CC cancer and treating ovarian cancer. This sequence represents the protein
CC of an ovarian specific gene of the invention.
XX
XX
XX Sequence 109 AA;
SQ
Query Match 100.0%; Score 104; DB 8; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SSVLVGYSPPNNEPLTGS 20
DB 31 SSVLVGYSPPNNEPLTGS 50
RESULT 6
ADM30756
ID ADM30756 standard; protein; 174 AA.
XX
XX ADM30756;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human cancer linked protein referred to as 192292 SeqID 4.
XX
XX human; cancer-related gene; screening assay; immuno-conjugate;
KW cytostatic; immunostimulant; cancer; ovarian cancer.
XX
XX Homo sapiens.
XX
XX WO2003075854-A2.
XX
XX 18-SEP-2003.
XX
XX

This invention relates to a novel method of identifying an agent that modulates the activity of a cancer-related gene. Specifically, it refers to a screening assay to identify potential antitumor agents, as well as methods to assess the cancerous state of a cell. The present invention describes antibodies against the expression products of these cancer-related genes that are capable of targeting a cancerous cell in vivo. Furthermore it provides immuno-conjugates containing such antibodies, which can be used to deliver target therapeutics thereto. Accordingly, these compositions that exhibit cytostatic and immunostimulant activities are useful for preventing or treating cancer in humans and in particular the treatment of ovarian cancer. This polypeptide sequence is a protein encoded by a human cancer-linked gene of the invention.

QY 1 SSVLDGYSPPNRNEPLTGNS 20
 |||||
Db 96 SSVLDGYSPPNRNEPLTGNS 115

DT	12-MAR-2003	(first entry)
XX		
AC	ABU54859;	
...		

Human; CA125; protein repeat; chromosome 19q13.2; amino terminal domain, amino terminal extension; carboxy terminal domain; vaccine; cancer; ovarian cancer; carcinoma.

PN WO200283866-A2.

12-APR-2002: 2002WO-IIS011734

PR 19-JUN-2001; 2001US-0299380P.
PR 27-SEP-2001; 2001US-00665738

PA (UYAR-) UNIV ARKANSAS.

PI O'brien T, Beard J, Underwood L;

New CA125 molecules, useful as a gold standard for detecting and monitoring the presence of CA125 antigen which can be used for diagnosing, monitoring or treating patients with cancer or for developing vaccine against cancer.

83

The invention relates to a CA125 protein comprising: (a) an extracellular amino terminal domain; (b) an amino terminal extension; (c) a multiple repeat domain; and (d) a carboxy terminal domain. The extracellular amino terminal domain comprises 5 genomic exons, the amino terminal extension comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and the carboxy terminal domain comprises a transmembrane anchor with a short cytoplasmic domain, and further comprises 9 genomic exons. The gene for CA125 is located on human chromosome 19q 13.2. Also included are isolated CA125 repeat domains, nucleic acids (including variants, homologues and degenerate versions) encoding CA125 proteins or repeat units, a vector comprising the nucleic acid, a cultured cell comprising the vector, a method of expressing CA125 antigen in a cell, the amino acid sequences of the CA125 repeat units (or their variants, fragments or sequences 50% identical to them), a purified antibody that selectively binds to an epitope in the receptor-binding domain of CA125 protein, a diagnostic for detecting and monitoring the presence of CA125 antigen (comprising recombinant CA125 having at least one repeat unit of the CA125 repeat domain including epitope binding sites), a therapeutic vaccine to treat mammals with elevated CA125 antigen levels or at risk of developing a disease or disease recurrence associated with elevated CA125 antigen levels (comprising recombinant CA125 repeat domains including epitope binding sites) and an antisense oligonucleotide that inhibits the expression of CA125. The CA125 molecule, particularly the multiple repeat domains are useful as a gold standard for detecting and monitoring the presence of CA125 antigen, which can be used for diagnosing, monitoring or treating patients with ovarian cancer and other carcinomas where CA125 is expressed. The molecules are also useful for developing a vaccine against cancer. The present sequence is a CA125 repeat protein

Sequence 284 AA:

Query Match	100.0%	Score 104;	DB 6;	Length 284;
Best Local Similarity	100.0%;	Pred. No. 5,66-08;		
Matches 20; Conservative	0;	Mismatches	0;	Gaps 0;

Qy 1 SSVLDGYSPPNRNEPTIGNS 20
Db 206 SSVLDGYSPPNRNEPTIGNS 225

RESULT 8
AAE12623
ID AAE12623 standard; protein; 438 AA.

AC AAE12623;

DT 03-JAN-2002 (first entry)

DE Human gene 1 encoded secreted protein HIPA230.

KM Human, ovarian cancer antigen; proliferative disorder; cancer; tumour;
KM respiratory system disorder; asthma; haematopoietic disorder; skin aging
KM immune system disorder; AIDS; skin disorder; autoimmune disease; allergy
KM Rheumatoid arthritis; inflammation; neurological disorder; septic shock;
KM Alzheimer's disease; Parkinson's disease; diabetes; angiocenic disorder;
KM atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
KM epithelial cell proliferation; transplantation; chemotaxis; infection;
KM food additive; wound healing; endocrine disorder; kidney disorder;
KM gene therapy; cytostatic.

Homo sapiens.

PN WO200170804-A1.


```
XX 27-SEP-2001.
PD
XX
XX 16-MAR-2001; 2001WO-US008585.
PF
XX 17-MAR-2000; 2000US-0190076P.
XX 23-AUG-2000; 2000US-0227009P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Birse CE, Rosen CA;
XX WPI; 2001-639119/73.
XX DR N-PSDB; AAD20618.
XX
XX Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
PT metastases.
XX
XX Claim 11; Page 412-414; 427pp; English.
XX
XX AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
CC ovarian cancer associated protein (collectively known as ovarian cancer
CC antigen) genes, and AAE12623-AAE12629 represent the proteins they
CC encode. AAE12630-AAE12638 represent human ovarian cancer antigenic
CC fragments or variants. Ovarian cancer antigens and their corresponding
CC DNAs are used in the prevention, diagnosis and treatment of diseases
CC associated with their inappropriate expression. These disorders include
CC proliferative disorders, cancer, tumours, respiratory system disorders,
CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
CC inflammation, allergies, neurological disorders (e.g., Alzheimer's
CC disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, endocrine disorders and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, to identify their cognate ligands or binding
CC partners, in chemotaxis and can be used as a food additive. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above and in diagnostic
CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
CC the invention is used in gene therapy. The present sequence represents a
CC human ovarian cancer antigen of the invention
XX
XX Sequence 438 AA:
SQ
XX
XX Query Match 100.0%; Score 104; DB 4; Length 438;
XX Best Local Similarity 100.0%; Pred. No. 9.2e-08;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 SSVLVGYSPPNREPLTGN 20
QY |||||
DB 360 SSVLVGYSPPNREPLTGN 379
XX
XX RESULT 9
XX ABP30973
ID ABP30973 standard; protein; 438 AA.
XX
XX ABP30973;
XX
XX 02-JUL-2002 (first entry)
XX
XX Partial protein sequence of clone 0772P partial cDNA sequence 4.
XX
XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
XX Homo sapiens.
XX
XX WO200206317-A2.
XX
XX 24-JAN-2002.
XX
XX
```

```
XX 17-JUL-2001; 2001WO-US022635.
PF
XX
XX 17-JUL-2000; 2000US-00617747.
XX 10-AUG-2000; 2000US-00636801.
XX 20-SEP-2000; 2000US-00667857.
XX 04-APR-2001; 2001US-00827271.
XX 18-JUN-2001; 2001US-00884441.
XX
XX (CORI-) CORIYA CORP.
XX
XX Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
XX PI Reed SG, Vedvick TS, Carter D, Hill P, Albone B;
XX WPI; 2002-164781/21.
XX DR N-PSDB; ABN72979.
XX
XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
XX Example 13; Page 357; 408pp; English.
XX
XX This invention relates to polypeptides comprising an immunogenic portion
CC of an ovarian carcinoma protein which acts as an immunostimulant and is
CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations and antigen presenting cells that express
CC the polypeptides are useful for stimulating an immune response in a
CC patient and treating ovarian cancer. This sequence represents protein
CC related to the invention
XX
XX Sequence 438 AA:
SQ
XX
XX Query Match 100.0%; Score 104; DB 5; Length 438;
XX Best Local Similarity 100.0%; Pred. No. 9.2e-08;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 SSVLVGYSPPNREPLTGN 20
QY |||||
DB 360 SSVLVGYSPPNREPLTGN 379
XX
XX RESULT 10
XX ADA08636
ID ADA08636 standard; protein; 438 AA.
XX
XX ADA08636;
XX
XX 06-NOV-2003 (first entry)
XX
XX Human 0772P partial protein #6.
XX
XX human; gene therapy; ovarian cancer; cancer.
XX
XX Homo sapiens.
XX
XX US2003091580-A1.
XX
XX 15-MAY-2003.
XX
XX 17-JUL-2001; 2001US-00907969.
XX
XX 18-JUN-2001; 2001US-00884441.
XX
XX (MITC/) MITCHAM J L.
XX (KING/) KING G E.
XX (ALGA/) ALGATE P A.
XX (FLIN/) FLING S P.
XX (RETT/) RETTER M W.
XX (FANG/) FANGER G R.
XX (REED/) REED S G.
XX (VEDV/) VEDVICK T S.
XX (CART/) CARTER D.
XX
```

PA (HILL/) HILL P.
 PA (ALBO/) ALBONE E.
 XX
 PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
 PI Reed SG, Vedrick TS, Carter D, Hill P, Albone E;
 XX
 DR WPI: 2003-532352/50.
 XX
 PT New isolated 0772p polypeptides and polynucleotides, useful in gene
 PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
 PT cancer.
 XX
 PS Example 13; SEQ ID NO 483; 371pp; English.
 CC
 CC The invention relates to an isolated 0772p polypeptide, which has the
 CC structure fully defined in the specification. The composition containing
 CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
 CC or antigen presenting cells are useful for stimulating an immune response
 CC and treating ovarian cancer. Detecting the presence of the
 CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
 CC carcinoma cDNAs and protein cDNAs were identified using microarray
 CC technology. The present sequence represents a human ovarian carcinoma
 CC antigen.
 CC
 SQ Sequence 438 AA;

Query Match 100.0%; Score 104; DB 7; Length 438;
 Best Local Similarity 100.0%; Pred. No. 9.2e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRNEPLTGNS 20
 |||||
 Db 360 SSVLVGYSPPNRNEPLTGNS 379

RESULT 11

ADP08979
 ID ADF08979 standard; protein; 438 AA.

AC ADF08979;
 XX

DT 12-FEB-2004 (first entry)
 XX

DE Secreted ovarian carcinoma antigen seqid 483.
 XX

XX gene therapy; protein therapy; vaccine; antibody inhibition;
 XX breast cancer; restorative therapy; diagnostic agent; immunoassay;
 KW secreted ovarian carcinoma antigen.

OS Homo sapiens.
 XX

PN US2003124140-A1.
 XX

PD 03-JUL-2003.
 XX

PF 17-JUL-2002; 2002US-00198053.
 XX

PR 17-DEC-1998; 98US-00215681.
 XX

PR 17-DEC-1998; 98US-00216003.
 XX

PR 23-JUN-1999; 99US-00338933.
 XX

PR 24-SEP-1999; 99US-00404879.
 XX

PR 17-JUL-2000; 2000US-00617747.
 XX

PR 10-AUG-2000; 2000US-00636801.
 XX

PR 20-SEP-2000; 2000US-00667857.
 XX

PR 04-APR-2001; 2001US-00827271.
 XX

PR 18-JUN-2001; 2001US-00884441.
 XX

PR 17-JUL-2001; 2001US-00907969.
 XX

PA (CORI-) CORIXA CORP.
 PA
 XX Bangur CS, Retter MW, Fanger GR, Hill P;
 PI
 XX WPI: 2003-897152/82.
 DR
 XX

XX
 PT Oncogenic nucleic acids useful for the prevention, diagnosis and
 PT treatment of breast cancer.
 XX
 PS Example 13; SEQ ID NO 483; 399pp; English.
 XX

CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancer. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patient's own production of (II).
 CC Additionally, (I) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigen in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and
 CC antagonists may be used to regulate expression and activity and as
 CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunoassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.
 CC
 SQ Sequence 438 AA;

Query Match 100.0%; Score 104; DB 7; Length 438;
 Best Local Similarity 100.0%; Pred. No. 9.2e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRNEPLTGNS 20
 |||||
 Db 360 SSVLVGYSPPNRNEPLTGNS 379

RESULT 12

ABU54709
 ID ABU54709 standard; protein; 439 AA.

AC ABU54709;
 XX

DT 12-MAR-2003 (first entry)
 XX

DE Human CA125 protein carboxy terminus.
 XX

XX Human; CA125; protein repeat; chromosome 19q 13.2; amino terminal domain;
 KW amino terminal extension; carboxy terminal domain; vaccine; cancer;
 KW ovarian cancer; carcinoma.
 XX

OS Homo sapiens.
 XX

PN WO200283866-A2.
 XX

PD 24-OCT-2002.
 XX

PF 12-APR-2002; 2002WO-US011734.
 XX

PR 17-APR-2001; 2001US-0284175P.
 XX

PR 19-JUN-2001; 2001US-0299380P.
 XX

PR 27-SEP-2001; 2001US-00965738.
 XX

PR 21-DEC-2001; 2001US-0345180P.
 XX

PA (UVAR-) UNIV ARKANSAS.
 XX

PI O'Brien T, Beard J, Underwood L;
 XX

DR WPI: 2003-093013/08.
 XX

DR N-PSDB; ABX72616.
 XX

PT New CA125 molecules, useful as a gold standard for detecting and
 monitoring the presence of CA125 antigen which can be used for

PT diagnosing, monitoring or treating patients with cancer or for developing
PT vaccine against cancer.

PS Claim 23; Page 111; 694pp; English.

XX The invention relates to a CA125 protein comprising: (a) an extracellular
CC amino terminal domain; (b) an amino terminal extension; (c) a multiple
CC repeat domain; and (d) a carboxy terminal domain. The extracellular amino
CC terminal domain comprises 5 genomic exons, the amino terminal extension
CC comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and
CC the carboxy terminal domain comprises a transmembrane anchor with a short
CC cytoplasmic domain, and further comprises 9 genomic exons. The gene for
CC CA125 is located on human chromosome 19q 13.2. Also included are isolated
CC CA125 repeat domains, nucleic acids (including variants, homologues and
CC degenerate versions) encoding CA125 proteins or repeat units, a vector
CC comprising the nucleic acid, a cultured cell comprising the vector, a
CC method of expressing CA125 antigen in a cell, the amino acid sequences of
CC the CA125 repeat units (or their variants, fragments or sequences 50%
CC identical to them), a purified antibody that selectively binds to an
CC epitope in the receptor-binding domain of CA125 protein, a diagnostic for
CC detecting and monitoring the presence of CA125 antigen (comprising
CC recombinant CA125 having at least one repeat unit of the CA125 repeat
CC domain including epitope binding sites), a therapeutic vaccine to treat
CC mammals with elevated CA125 antigen levels or at risk of developing a
CC disease or disease recurrence associated with elevated CA125 antigen
CC levels (comprising recombinant CA125 repeat domains including epitope
CC binding sites) and an antisense oligonucleotide that inhibits the
CC expression of CA125. The CA125 molecule, particularly the multiple repeat
CC domains are useful as a gold standard for detecting and monitoring the
CC presence of CA125 antigen, which can be used for diagnosing, monitoring
CC or treating patients with ovarian cancer and other carcinomas where CA125
CC is expressed. The molecules are also useful for developing a vaccine
CC against cancer. The present sequence is a CA125 repeat protein

XX Sequence 439 AA;

Query Match 100.0%; Score 104; DB 6; Length 439;
Best Local Similarity 100.0%; Pred. No. 9,2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLDVGSPPNRPPLTGNS 20
|||
DB 361 SSVLDVGSPPNRPPLTGNS 380

RESULT 13

AAE12634
ID AAE12634 standard; protein; 526 AA.

AC AAE12634;
XX

DT 03-JAN-2002 (first entry)

DE Human gene 1 encoded secreted protein fragment, SEQ ID NO: 30.

XX Human; ovarian cancer antigen; proliferative disorder; cancer; tumour;
XX respiratory system disorder; asthma; haematopoietic disorder; skin aging;
XX immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
XX Rheumatoid arthritis; inflammation; neurological disorder; septic shock;
XX Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
XX atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
XX epithelial cell proliferation; transplantation; chemotaxis; infection;
XX food additive; wound healing; endocrine disorder; kidney disorder;
XX gene therapy; cytostatic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX MISC-difference 116

XX /label= Unknown

XX /note="xaa equals any of the naturally occurring L-amino
XX acids"

PN WO200170804-A1.

XX 27-SEP-2001.

XX 16-MAR-2001; 2001WO-US008585.

XX 17-MAR-2000; 2000US-0190076P.

XX 23-AUG-2000; 2000US-0227009P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2001-639119/73.

PT Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
PT metastases.

PS Disclosure; Page 10; 427pp; English.

XX AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
CC ovarian cancer associated protein (collectively known as ovarian cancer
CC antigen) genes, and AAE12623-AAE12629 represent the proteins they
CC encode. AAE12630-AAE12638 represent human ovarian cancer antigenic
CC fragments or variants. Ovarian cancer antigens and their corresponding
CC DNAs are used in the prevention, diagnosis and treatment of diseases
CC associated with their inappropriate expression. These disorders include
CC proliferative disorders, cancer, tumours, respiratory system disorders,
CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
CC skin disorders, allergic, neurological disorders (e.g., Rheumatoid arthritis),
CC inflammation, Parkinson's disease, septic shock, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, endocrine disorders and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, to identify their cognate ligands or binding
CC partners, in chemotaxis and can be used as a food additive. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above and in diagnostic
CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
CC the invention is used in gene therapy. The present sequence represents a
CC human ovarian cancer antigenic fragment of the invention

XX Sequence 526 AA;

Query Match 100.0%; Score 104; DB 4; Length 526;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLDVGSPPNRPPLTGNS 20
|||
DB 448 SSVLDVGSPPNRPPLTGNS 467

RESULT 14

ID ADS94302 standard; protein; 748 AA.

AC ADS94302;

DT 02-DEC-2004 (first entry)

XX CA 125/0772P 3-repeat amino acid sequence SEQ ID NO:1.

XX antibody; antigen-binding antibody fragment;

XX cell-associated CA 125/0772P; monoclonal antibody; cytostatic;

XX immunostimulant; mediator of lysis; tumour; cell proliferative disorder;

XX cancer; cervical cancer; uterine cancer; breast cancer; lung cancer;

XX ovarian cancer.

XX Synthetic.

XX WO2004035537-A2.
 XX
 XX 29-APR-2004.
 XX
 XX 15-OCT-2003; 2003WO-US032945.
 XX
 XX 16-OCT-2002; 2002US-0418828P.
 XX 10-JUL-2003; 2003US-0485986P.
 XX
 XX (EURO-) EUROCELLTIGUE SA.
 XX
 XX Albione EF, Solitis DA;
 XX
 XX WPI; 2004-357171/33.
 XX
 XX Novel isolated antibody, or antigen-binding antibody fragment binding
 XX with cell-associated CA 125/O772P polypeptide relative to shed CA
 XX 125/O772 polypeptide, useful for ameliorating cervical or ovarian cancer.
 XX
 XX Example; SEQ ID NO 1, 153bp; English.

The present invention describes an isolated antibody, or an antigen-binding antibody fragment (I), that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772 polypeptide. Also described: (1) a monoclonal antibody (II) produced by hybridoma 4E7; (2) a monoclonal antibody that competes with binding of (II); (3) a hybridoma as deposited in (II); (4) an isolated nucleic acid molecule (III) comprising a nucleotide sequence that encodes a variable chain region of (II); (5) a pharmaceutical composition comprising an antibody or an antigen-binding antibody fragment that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a carrier; (6) a pharmaceutical composition comprising a monoclonal antibody or an antigen-binding monoclonal antibody fragment that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a carrier; (7) an article of manufacture (IV) comprising packaging material and a composition comprising an antibody, or an antigen-binding antibody fragment that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a carrier; (8) a fusion polypeptide (V) comprising an antibody, or an antigen-binding antibody fragment, which preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a heterologous agent; (9) ameliorating (M1) a symptom of a CA 125/O772P-related disorder; (10) a monoclonal antibody (VI) chosen from 325.1, 621.1, 633.1, 654.1, 725.1, 8G9, 7F10, 8A1, 8AC3, 15C9, 8E3, 8B5, 7G10, 16C7, 7C6, 7H1, 16H9, 7A11, 4E7, 117.1, 368.1, 446.1, 501.1, and 776.1 or its antigen-binding antibody fragment; (11) an antibody or antigen binding antibody fragment that competes with (VI); and (12) a pharmaceutical composition comprising (VI) and a carrier. (I) has cytostatic activity, and can be used as an immunostimulant and a mediator of lysis of positive tumour cell. (I) is useful for ameliorating a symptom of a CA 125/O77P-related disorder which is a cell proliferative disorder such as cancer, cervical or uterine cancer, breast or lung cancer or ovarian cancer. (V) is useful diagnostically for monitoring the development or progression of cancer or tumour as part of clinical testing procedure. The present sequence represents the CA 125/O772P 3-repeat amino acid sequence, which is used in the exemplification of the present invention.

XX Sequence 748 AA;

Query Match 100.0%; Score 104; DB 8; Length 748;

Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVDSPPNNEPLTGS 20
 |||||
 DB 691 SSVLVDSPPNNEPLTGS 710

RESULT 15
 ADS94303

ID ADS94303 standard; protein; 809 AA.
 XX
 XX ADS94303;
 XX
 XX 02-DEC-2004 (first entry)
 XX
 XX CA 125/O772P 3-repeat TW amino acid sequence SEQ ID NO:2.
 XX
 XX antibody; antigen-binding antibody fragment;
 XX cell-associated CA 125/O772P; monoclonal antibody; cytostatic;
 XX immunostimulant; mediator of lysis; tumour; cell proliferative disorder;
 XX cancer; cervical cancer; uterine cancer; breast cancer; lung cancer;
 XX ovarian cancer.
 XX
 XX Synthetic.
 XX
 XX WO2004035537-A2.
 XX
 XX 29-APR-2004.
 XX
 XX 15-OCT-2003; 2003WO-US032945.
 XX
 XX 16-OCT-2002; 2002US-0418828P.
 XX 10-JUL-2003; 2003US-0485986P.
 XX
 XX (EURO-) EUROCELLTIGUE SA.
 XX
 XX Albione EF, Solitis DA;
 XX
 XX WPI; 2004-357171/33.
 XX
 XX Novel isolated antibody, or antigen-binding antibody fragment binding
 XX with cell-associated CA 125/O772P polypeptide relative to shed CA
 XX 125/O772 polypeptide, useful for ameliorating cervical or ovarian cancer.
 XX
 XX Example; SEQ ID NO 2; 153bp; English.

The present invention describes an isolated antibody, or an antigen-binding antibody fragment (I), that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772 polypeptide. Also described: (1) a monoclonal antibody (II) produced by hybridoma 4E7; (2) a monoclonal antibody that competes with binding of (II); (3) a hybridoma as deposited in (II); (4) an isolated nucleic acid molecule (III) comprising a nucleotide sequence that encodes a variable chain region of (II); (5) a pharmaceutical composition comprising an antibody or an antigen-binding antibody fragment that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a carrier; (6) a pharmaceutical composition comprising a monoclonal antibody or an antigen-binding monoclonal antibody fragment that preferentially binds cell-associated CA 125/O772P polypeptide relative to shed CA 125/O772P polypeptide, and a heterologous agent; (9) ameliorating (M1) a symptom of a CA 125/O772P-related disorder; (10) a monoclonal antibody (VI) chosen from 325.1, 621.1, 633.1, 654.1, 725.1, 8G9, 7F10, 8A1, 8AC3, 15C9, 8E3, 8B5, 7G10, 16C7, 7C6, 7H1, 16H9, 7A11, 4E7, 117.1, 368.1, 446.1, 501.1, and 776.1 or its antigen-binding antibody fragment; (11) an antibody or antigen binding antibody fragment that competes with (VI); and (12) a pharmaceutical composition comprising (VI) and a carrier. (I) has cytostatic activity, and can be used as an immunostimulant and a mediator of lysis of positive tumour cell. (I) is useful for ameliorating a symptom of a CA 125/O77P-related disorder which is a cell proliferative disorder such as cancer, cervical or uterine cancer, breast or lung cancer or ovarian cancer. (V) is useful diagnostically for monitoring the development or progression of cancer or tumour as part of clinical testing procedure. The present sequence represents the CA 125/O772P 3-repeat TW amino acid sequence, which is

CC used in the exemplification of the present invention.

XX Sequence 809 AA;

Query Match 100.0%; Score 104; DB 8; Length 809;
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVNDGYSPPNREPLTGNS 20
 DB 691 SSVLVNDGYSPPNREPLTGNS 710

Search completed: October 27, 2005, 18:44:23
 Job time : 73.2093 secs

This page blank (us9:9)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using ew model

Run on: October 27, 2005, 17:03:42 ; Search time 17.3953 Seconds
(without alignments)
85.826 Million cell updates/sec

Title: US-10-612-090-14
Perfect score: 104
Sequence: 1 SSVLDVGYSPNNEPLTGS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents_AA:*
2: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
7: /cgn2_6/prodata/1/1aa/Backfillseq1.pep:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	94.2	438	4	US-09-404-879A-390 Sequence 390, App
2	98	94.2	438	4	US-09-667-857-390 Sequence 390, App
3	98	94.2	833	4	US-09-404-879A-389 Sequence 389, App
4	98	94.2	833	4	US-09-667-857-389 Sequence 389, App
5	98	94.2	914	4	US-09-404-879A-312 Sequence 312, App
6	98	94.2	914	4	US-09-338-933-312 Sequence 312, App
7	98	94.2	914	4	US-09-667-857-312 Sequence 312, App
8	50	48.1	3519	3	US-09-428-517-4 Sequence 4, Appl1
9	47	45.2	2142	4	US-09-540-236-3459 Sequence 3459, Ap
10	45	43.3	513	1	US-08-464-340A-2 Sequence 2, Appl1
11	45	43.3	513	4	US-09-949-016-10288 Sequence 10288, A
12	45	43.3	513	5	PCT-US94-0849A-2 Sequence 2, Appl1
13	45	43.3	811	4	US-09-489-039A-13839 Sequence 13839, A
14	45	43.3	1048	4	US-09-171-699-10 Sequence 10, Appl1
15	44	42.3	1715	4	US-09-696-115B-17 Sequence 17, Appl1
16	43	41.8	299	4	US-09-502-540-13589 Sequence 13589, A
17	43	41.3	391	2	US-08-759-581B-13 Sequence 13, Appl1
18	43	41.3	391	3	US-09-304-711-13 Sequence 13, Appl1
19	43	41.3	391	4	US-09-173-281-13 Sequence 13, Appl1
20	43	41.3	504	4	US-09-489-039A-11966 Sequence 11966, A
21	43	41.3	575	4	US-09-489-039A-8630 Sequence 8630, Ap
22	43	41.3	994	4	US-09-515-311-13 Sequence 13, Appl1
23	43	41.3	1346	3	US-09-320-878-4 Sequence 4, Appl1
24	43	41.3	1346	3	US-09-105-537-37 Sequence 37, Appl1
25	43	41.3	1346	4	US-09-141-908-5 Sequence 5, Appl1
26	43	41.3	1346	4	US-09-657-440-4 Sequence 4, Appl1
27	43	41.3	11877	3	US-09-105-537-6 Sequence 6, Appl1

28	42.5	40.9	428	4	US-09-538-092-531 Sequence 531, App
29	42	40.4	328	4	US-09-902-540-10796 Sequence 10796, A
30	42	40.4	422	4	US-09-831-630-10 Sequence 10, Appl1
31	42	40.4	721	4	US-09-328-352-7781 Sequence 7781, Ap
32	41.5	39.9	218	4	US-09-328-352-6869 Sequence 6869, Ap
33	41	39.4	201	3	US-09-015-734-12 Sequence 12, Appl1
34	41	39.4	201	4	US-09-515-311-12 Sequence 12, Appl1
35	41	39.4	213	4	US-09-489-039A-14310 Sequence 14310, A
36	41	39.4	236	3	US-09-015-734-7 Sequence 7, Appl1
37	41	39.4	236	4	US-09-515-311-7 Sequence 7, Appl1
38	41	39.4	255	3	US-09-015-734-2 Sequence 2, Appl1
39	41	39.4	255	4	US-09-515-311-2 Sequence 2, Appl1
40	41	39.4	468	4	US-09-248-796A-27314 Sequence 27314, A
41	41	39.4	641	4	US-09-252-991A-26329 Sequence 26329, A
42	41	39.4	655	4	US-09-248-796A-14460 Sequence 14460, A
43	41	39.4	746	4	US-09-248-796A-19979 Sequence 19979, A
44	41	39.4	922	4	US-09-489-039A-8938 Sequence 8938, Ap
45	41	39.4	929	4	US-09-328-352-4591 Sequence 4591, Ap

ALIGNMENTS

RESULT 1
US-09-404-879A-390
; Sequence 390, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-390
Query Match 94.2%; Score 98; DB 4; Length 438;
Best Local Similarity 95.0%; Pred. No. 1.8e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 SSVLDVGYSPNNEPLTGS 20
DB 360 SSVLDVGYSPNNEPLTGS 379
RESULT 2
US-09-667-857-390
; Sequence 390, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455

SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-667-857-390

Query Match 94.2%; Score 98; DB 4; Length 438;
Best Local Similarity 95.0%; Pred. No. 1.8e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSVLDGYSPPNNEPLTGS 20
DB 360 SSVLDGTFPNNNEPLTGS 379

RESULT 3
US-09-404-879A-389
; Sequence 389, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-404-879A-389

Query Match 94.2%; Score 98; DB 4; Length 833;
Best Local Similarity 95.0%; Pred. No. 3.8e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSVLDGYSPPNNEPLTGS 20
DB 755 SSVLDGTFPNNNEPLTGS 774

RESULT 4
US-09-667-857-389
; Sequence 389, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Reller, Steven G.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-667-857-389

Query Match 94.2%; Score 98; DB 4; Length 833;
Best Local Similarity 95.0%; Pred. No. 3.8e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSVLDGYSPPNNEPLTGS 20
DB 755 SSVLDGTFPNNNEPLTGS 774

RESULT 5
US-09-404-879A-312
; Sequence 312, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRF
; ORGANISM: Homo sapien
US-09-404-879A-312

Query Match 94.2%; Score 98; DB 4; Length 914;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSVLDGYSPPNNEPLTGS 20
DB 836 SSVLDGTFPNNNEPLTGS 855

RESULT 6
US-09-338-933-312
; Sequence 312, Application US/09338933
; Patent No. 6468931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRF
; ORGANISM: Homo sapien
US-09-338-933-312

Query Match 94.2%; Score 98; DB 4; Length 914;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SSVLDGYSPPNNEPLTGS 20
DB 836 SSVLDGTFPNNNEPLTGS 855

RESULT 7
US-09-667-857-312
; Sequence 312, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Rling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: PastsEQ for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-667-857-312

```

```

Query Match          94.2%; Score 98; DB 4; Length 914;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 SSVLVGYSPPNREPLTGN 20
DB      836 SSVLVGYSPPNREPLTGN 855

```

```

RESULT 8
US-09-428-517-4
; Sequence 4, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Beljach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3519
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
US-09-428-517-4

```

```

Query Match          48.1%; Score 50; DB 3; Length 3519;
Best Local Similarity 62.5%; Pred. No. 87;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      3 VLVGYSPPNREPLTG 18
DB      3401 VLVVYPPGROEPVFG 3416

```

```

RESULT 9
US-09-540-236-3459
; Sequence 3459, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:

```

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3459
; LENGTH: 2142
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-3459

```

```

Query Match          45.2%; Score 47; DB 4; Length 2142;
Best Local Similarity 47.4%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      2 SSVLVGYSPPNREPLTGN 19
DB      105 STAIGGKPPNRPQDNN 122

```

```

RESULT 10
US-08-464-340A-2
; Sequence 2, Application US/08464340A
; Patent No. 5710019
; GENERAL INFORMATION:
; APPLICANT: LI, ET AL.
; TITLE OF INVENTION: Human Potassium Channel 1 and 2 Proteins
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,340A
; FILING DATE: June 5, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08449
; FILING DATE: 28 JUL 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-415
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 AMINO ACIDS
; TYPE: AMINO ACIDS
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-464-340A-2

```

```

Query Match          43.3%; Score 45; DB 1; Length 513;
Best Local Similarity 47.4%; Pred. No. 57;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```

```

QY      2 SSVLVGYSPPNREPLTGN 20
DB      315 TLVVDGAAGRRKPGAGNS 333

```

```

RESULT 11
US-09-949-016-10288
; Sequence 10288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01107
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10288
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10288

Query Match          43.3% Score 45; DB 4; Length 513;
Best Local Similarity 47.4% Pred. No. 57;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY      2 SVLVGYSPPNRNEPLTGN 20
Db      315 TLVVDGAAAGRRKPGAGNS 333

PCT-US94-08449A-2
; Sequence 2, Application PC/TUS9408449A
; GENERAL INFORMATION:
; APPLICANT: LI, ET AL.
; TITLE OF INVENTION: Potassium Channel Protein 1 and 2
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08449A
; FILING DATE: SUBMITTED HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; TELEPHONE: 201-994-1744
; TELEFAX: 201-994-1700
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 AMINO ACIDS

```

```

; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
PCT-US94-08449A-2

Query Match          43.3% Score 45; DB 5; Length 513;
Best Local Similarity 47.4% Pred. No. 57;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY      2 SVLVGYSPPNRNEPLTGN 20
Db      315 TLVVDGAAAGRRKPGAGNS 333

RESULT 13
US-09-489-039A-13839
; Sequence 13839, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13839
; LENGTH: 811
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13839

Query Match          43.3% Score 45; DB 4; Length 811;
Best Local Similarity 66.7% Pred. No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY      8 YSPNRNEPLTGN 19
Db      750 YSPNINIPINGN 761

RESULT 14
US-09-171-699-10
; Sequence 10, Application US/09171699
; Patent No. 6448389
; GENERAL INFORMATION:
; APPLICANT: The Wistar Institute of, Anatomy & Biology
; ADDRESSEE: Gonczol, Eva
; ADDRESSEE: Berencsi, Klara
; Kari, Csaba
; TITLE OF INVENTION: NO. 6448389el Cyromegalovirus DNA Constructs and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,699
; FILING DATE: 19-Jan-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 60/015,717
FILING DATE: 23-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kodroff, Cathy A.
REGISTRATION NUMBER: 33,980
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1048 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-171-699-10

Query Match 43.3%; Score 45; DB 4; Length 1048;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 DGSPNNRNP 15
DB 537 DGYPNRDP 546

RESULT 15
US-09-696-115B-17
Sequence 17, Application US/09696115B
Patent No. 6660497
GENERAL INFORMATION:
APPLICANT: Bulla, Lee
APPLICANT: Candae, Mehmet
TITLE OF INVENTION: Pectinophora gossypiella (Pink Bollworm) Bacillus thuringiensis
FILE REFERENCE: 52418-20003.00
CURRENT APPLICATION NUMBER: US/09/696,115B
CURRENT FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/161,564
PRIOR FILING DATE: 1999-10-26
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin version 3.0
SEQ ID NO 17
LENGTH: 1715
TYPE: PRT
ORGANISM: B. mori
US-09-696-115B-17

Query Match 42.3%; Score 44; DB 4; Length 1715;
Best Local Similarity 61.5%; Pred. No. 3.4e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 YSPNNEPLTGN 20
DB 1028 YAPDRDEPDND 1040

Search completed: October 27, 2005, 17:34:31
Job time: 18.3953 secs

This page Blank (1894)

Result No.	Score	Query Match	Length	DB	ID	Description
1	104	100.0	20	15	US-10-612-090-14	Sequence 14, App1
2	104	100.0	108	15	US-10-612-090-2	Sequence 2, App1
3	104	100.0	108	15	US-10-612-090-13	Sequence 13, App1
4	104	100.0	174	15	US-10-383-368-4	Sequence 4, App1
5	104	100.0	284	15	US-09-965-728-300	Sequence 300, App
6	104	100.0	438	9	US-09-884-441-483	Sequence 483, App
7	104	100.0	438	10	US-09-907-969-483	Sequence 483, App
8	104	100.0	438	14	US-10-198-053-483	Sequence 483, App
9	104	100.0	438	15	US-10-333-990-19	Sequence 19, App1
10	104	100.0	438	17	US-10-860-790-483	Sequence 483, App
11	104	100.0	439	10	US-09-965-738-148	Sequence 148, App

45	104	100.0	1890	20	US-11-050-926-217	Sequence 217, App
44	104	100.0	1890	15	US-10-253-286-314	Sequence 314, App
43	104	100.0	1890	15	US-10-245-871-314	Sequence 314, App
42	104	100.0	1890	14	US-10-097-340-217	Sequence 217, App
41	104	100.0	1889	14	US-10-142-515-5	Sequence 5, Appl1
40	104	100.0	1156	17	US-10-860-799-459	Sequence 459, App
39	104	100.0	1156	14	US-10-198-055-459	Sequence 459, App
38	104	100.0	1156	14	US-09-827-271-459	Sequence 459, App
37	104	100.0	1156	9	US-09-907-969-459	Sequence 459, App
36	104	100.0	1156	9	US-09-884-441-459	Sequence 459, App
35	104	100.0	1148	20	US-11-050-926-216	Sequence 216, App
34	104	100.0	1148	17	US-10-938-061-139	Sequence 129, App
33	104	100.0	1148	17	US-10-936-626-129	Sequence 129, App
32	104	100.0	1148	17	US-10-860-799-479	Sequence 479, App
31	104	100.0	1148	17	US-10-860-799-479	Sequence 479, App
30	104	100.0	1148	16	US-10-734-556-131	Sequence 131, App
29	104	100.0	1148	15	US-10-383-366-2	Sequence 2, Appl1
28	104	100.0	1148	15	US-10-257-021-56	Sequence 56, Appl1
27	104	100.0	1148	14	US-10-198-053-479	Sequence 479, App
26	104	100.0	1148	14	US-10-198-053-458	Sequence 458, App
25	104	100.0	1148	14	US-10-097-340-216	Sequence 216, App
24	104	100.0	1148	10	US-09-827-271-458	Sequence 458, App
23	104	100.0	1148	10	US-09-965-738-48	Sequence 48, Appl1
22	104	100.0	1148	10	US-09-932-419-2	Sequence 2, Appl1
21	104	100.0	1148	10	US-09-907-969-479	Sequence 479, App
20	104	100.0	1148	9	US-09-907-969-458	Sequence 458, App
19	104	100.0	1148	9	US-09-884-441-479	Sequence 479, App
18	104	100.0	1148	9	US-09-884-441-458	Sequence 458, App
17	104	100.0	809	17	US-10-687-035-2	Sequence 2, Appl1
16	104	100.0	748	17	US-10-687-035-1	Sequence 1, Appl1
15	104	100.0	563	14	US-10-142-515-4	Sequence 4, Appl1
14	104	100.0	545	14	US-10-243-243A-4	Sequence 4, Appl1
13	104	100.0	536	18	US-10-333-900-30	Sequence 30, App
12	104	100.0	456	18	US-10-858-412-226	Sequence 226, App

ALIGNMENTS

```

RESULT 1
US-10-612-090-14
; Sequence 14, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 14
; LENGTH: 20
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-612-090-14

```

```

Query Match          100.0%;   Score 104;  DB 15;   Length 20;
Best Local Similarity 100.0%;   Pred. No. 9.3e-10;
Matches 20;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0

OY          1  SSVLVVDGYSPPRNNEPLTGNS 20
            ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db          1  SSVLVVDGYSPPRNNEPLTGNS 20

RESULT 2
US-10-612-090-2
; Sequence 2, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:

```

```
/ APPLICANT: ImmunoGen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-612-090-2

Query Match          100.0%; Score 104; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 6,6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSVLDGYSPPNNEPLTGNS 20
DB      86 SSVLDGYSPPNNEPLTGNS 105

RESULT 3
US-10-612-090-13
/ Sequence 13, Application US/10612090
/ Publication No. US20040057952A1
/ GENERAL INFORMATION:
/ APPLICANT: ImmunoGen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 13
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion protein
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1)-(1)
/ OTHER INFORMATION: Glutathione S-transferase fusion site
US-10-612-090-13

Query Match          100.0%; Score 104; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 6,6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSVLDGYSPPNNEPLTGNS 20
DB      86 SSVLDGYSPPNNEPLTGNS 105

RESULT 4
US-10-383-368-4
/ Sequence 4, Application US/10383368
/ Publication No. US20040002091A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul B.
/ APPLICANT: Ebner, Reinhard
/ APPLICANT: Scrovel, Jeffrey W.
/ TITLE OF INVENTION: Cancer-Linked Gene as Target for Chemotherapy
/ FILE REFERENCE: 689290-123
/ CURRENT APPLICATION NUMBER: US/10/383,368
/ CURRENT FILING DATE: 2003-03-07
/ PRIOR APPLICATION NUMBER: US 60/362,527
/ PRIOR FILING DATE: 2002-03-07
```

```
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 4
/ LENGTH: 174
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-383-368-4

Query Match          100.0%; Score 104; DB 15; Length 174;
Best Local Similarity 100.0%; Pred. No. 1,1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSVLDGYSPPNNEPLTGNS 20
DB      96 SSVLDGYSPPNNEPLTGNS 115

RESULT 5
US-09-965-738-300
/ Sequence 300, Application US/09965738
/ Publication No. US2003014367A1
/ GENERAL INFORMATION:
/ APPLICANT: O'Brien, Timothy
/ TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic ar
/ FILE REFERENCE: 40715-258841
/ CURRENT APPLICATION NUMBER: US/09/965,738
/ CURRENT FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: US 60/284,175
/ PRIOR FILING DATE: 2001-04-17
/ NUMBER OF SEQ ID NOS: 306
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 300
/ LENGTH: 284
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-965-738-300

Query Match          100.0%; Score 104; DB 10; Length 284;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSVLDGYSPPNNEPLTGNS 20
DB      206 SSVLDGYSPPNNEPLTGNS 225

RESULT 6
US-09-884-441-483
/ Sequence 483, Application US/09884441
/ Patent No. US20020119158A1
/ GENERAL INFORMATION:
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Carter, Darrick
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121,462C7
/ CURRENT APPLICATION NUMBER: US/09/884,441
/ CURRENT FILING DATE: 2001-06-18
/ NUMBER OF SEQ ID NOS: 489
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 483
/ LENGTH: 438
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-884-441-483

Query Match          100.0%; Score 104; DB 9; Length 438;
Best Local Similarity 100.0%; Pred. No. 3,3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SSVLDGYSPPNNEPLTGNS 20
```


Db 360 SSVLDGYSPPNNEPLTGN 379

RESULT 7

US-09-907-969-483
; Sequence 483, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albome, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 483
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-969-483

Query Match 100.0%; Score 104; DB 10; Length 438;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSVLDGYSPPNNEPLTGN 20
|||||
Db 360 SSVLDGYSPPNNEPLTGN 379

RESULT 8
US-10-198-053-483
; Sequence 483, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 483
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-483

Query Match 100.0%; Score 104; DB 14; Length 438;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSVLDGYSPPNNEPLTGN 20
|||||
Db 360 SSVLDGYSPPNNEPLTGN 379

RESULT 9

US-10-333-900-19
; Sequence 19, Application US/10333900
; Publication No. US20040010121A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 7 Human Ovarian and Ovarian Cancer Associated Proteins
; FILE REFERENCE: PA006PCT
; CURRENT APPLICATION NUMBER: US/10/333,900
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/227,009
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/190,076
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-333-900-19

Query Match 100.0%; Score 104; DB 15; Length 438;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSVLDGYSPPNNEPLTGN 20
|||||
Db 360 SSVLDGYSPPNNEPLTGN 379

RESULT 10
US-10-860-790-483
; Sequence 483, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 483
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-483

Query Match 100.0%; Score 104; DB 17; Length 438;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SSVLDGYSPPNNEPLTGN 20
|||||
Db 360 SSVLDGYSPPNNEPLTGN 379

RESULT 11
US-09-965-738-148
; Sequence 148, Application US/09965738
; Publication No. US2003014367A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy
; TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic at
; FILE REFERENCE: 40715-258841
; CURRENT APPLICATION NUMBER: US/09/965,738
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/284,175

;; PRIOR FILING DATE: 2001-04-17
;; NUMBER OF SEQ ID NOS: 306
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 148
;; LENGTH: 439
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-965-738-148

Query Match 100.0%; Score 104; DB 10; Length 439;
Best Local Similarity 100.0%; Pred. No. 3,3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRPPLTGS 20
|||
Db 361 SSVLVGYSPPNRPPLTGS 380

RESULT 12
US-10-858-412-226
; Sequence 226, Application US/10858412
; Publication No. US20050214781A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Turner, Leah
; APPLICANT: Sun, Yongming
; TITLE OF INVENTION: Compositions, Splice Variants and Methods Relating to Ovarian Spe
; FILE REFERENCE: DEX-0488
; CURRENT APPLICATION NUMBER: US/10/858,412
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/474,854
; PRIOR FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 226
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-858-412-226

Query Match 100.0%; Score 104; DB 18; Length 456;
Best Local Similarity 100.0%; Pred. No. 3,5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRPPLTGS 20
|||
Db 378 SSVLVGYSPPNRPPLTGS 397

RESULT 13
US-10-333-900-30
; Sequence 30, Application US/10333900
; Publication No. US20040010121A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 7 Human Ovarian and Ovarian Cancer Associated Proteins
; FILE REFERENCE: PA006PCT
; CURRENT APPLICATION NUMBER: US/10/333,900
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/227,009
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 60/190,076
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (116)

;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-333-900-30

Query Match 100.0%; Score 104; DB 15; Length 526;
Best Local Similarity 100.0%; Pred. No. 4,1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRPPLTGS 20
|||
Db 448 SSVLVGYSPPNRPPLTGS 467

RESULT 14
US-10-243-243A-4
; Sequence 4, Application US/10243243A
; Publication No. US20030104442A1
; GENERAL INFORMATION:
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses Th
; FILE REFERENCE: 649-B
; CURRENT APPLICATION NUMBER: US/10/243,243A
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 10/142,515
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: PCT/US02/14768
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(545)
; OTHER INFORMATION: Deduced amino acid sequence of CA125/MUC16A (B4), Part Three
US-10-243-243A-4

Query Match 100.0%; Score 104; DB 14; Length 545;
Best Local Similarity 100.0%; Pred. No. 4,3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLVGYSPPNRPPLTGS 20
|||
Db 467 SSVLVGYSPPNRPPLTGS 486

RESULT 15
US-10-142-515-4
; Sequence 4, Application US/10142515
; Publication No. US20030078399A1
; GENERAL INFORMATION:
; APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses Th
; FILE REFERENCE: 649-A-US
; CURRENT APPLICATION NUMBER: US/10/142,515
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Human Being
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(583)

; OTHER INFORMATION: Deduced amino acid sequence of CA125/MUC16 (B4), Section three
US-10-142-515-4

Query Match 100.0%; Score 104; DB 14; Length 583;
Best Local Similarity 100.0%; Pred. No. 4.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SSVLYDGYSPNRPPLTGNS 20
|||
Db 505 SSVLYDGYSPNRPPLTGNS 524

Search completed: October 27, 2005, 18:31:45
Job time : 65 secs

SECRET

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 18:13:49 ; Search time 12.3721 Seconds
(without alignments)
155.538 Million cell updates/sec

Title: US-10-612-090-14

Perfect score: 104

Sequence: 1 SSVLVGVGSPNRRNEPLTGNLS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	50	48.1	3519	2	S43048	polyketide synthas
2	49	47.1	1666	1	C3GP	complement C3 prec
3	47	45.2	407	2	S49890	regulatory protein
4	46	44.2	1018	2	JC5799	alpha-D-mannosidas
5	46	44.2	1076	2	T26044	hypothetical prote
6	45	43.3	107	2	T46746	ubiquinol-cytochro
7	45	43.3	513	2	JC5920	potassium channel
8	45	43.3	715	2	AE1021	formate dehydrogen
9	45	43.3	1048	1	XPBR99	large structural p
10	45	43.3	1309	2	T39379	sexual differential
11	45	43.3	2214	2	T16305	hypothetical prote
12	44	42.3	186	2	C81903	hypothetical prote
13	44	42.3	253	2	D81118	conserved hypotet
14	44	42.3	445	2	AD2023	carboxyl-terminal
15	44	42.3	748	2	T35818	probable integral
16	44	42.3	1015	2	T32186	hypothetical prote
17	44	42.3	1355	2	T28715	hypothetical prote
18	44	42.3	1715	2	JE0128	Bombyx mori recept
19	44	41.8	1006	2	AD2195	hypothetical prote
20	43	41.3	367	2	T23983	hypothetical prote
21	43	41.3	426	1	F82315	conserved hypotet
22	43	41.3	715	1	DEECFS	formate dehydrogen
23	43	41.3	715	2	E91261	formate dehydrogen
24	43	41.3	715	2	A86102	formate dehydrogen
25	43	41.3	1346	2	T17412	polyketide synthas
26	42.5	40.9	428	2	S61823	hypothetical prote
27	42.5	40.9	1089	2	T21582	hypothetical prote
28	42	40.4	62	2	JQ0208	hypothetical 6.6K
29	42	40.4	62	2	S07166	hypothetical prote

30	42	40.4	62	2	S58538	hypothetical prote
31	42	40.4	70	2	T04410	hypothetical prote
32	42	40.4	78	2	C95976	hypothetical prote
33	42	40.4	295	2	S58850	homeotic protein a
34	42	40.4	320	2	T36057	hypothetical prote
35	42	40.4	474	2	C87556	hypothetical prote
36	42	40.4	507	1	A30828	steroid 17alpha-mo
37	42	40.4	582	2	S40176	Exor protein - Rhl
38	42	40.4	613	2	P81334	hypothetical prote
39	42	40.4	648	2	B40727	S-M checkpoint con
40	42	40.4	792	2	F83304	probable restricti
41	42	40.4	1663	1	C3RT	complement C3 prec
42	42	40.4	1778	2	T50074	probable nucleopor
43	41.5	39.9	225	2	A86903	hypothetical prote
44	41.5	39.9	735	2	G71203	probable translati
45	41	39.4	144	2	T29958	hypothetical prote

ALIGNMENTS

RESULT 1

S43048 polyketide synthase type I - Streptomyces antibioticus
N:Contains: acyl carrier protein; acyltransferase; ketosacylsynthase; ketoreductase; thio
C:Species: Streptomyces antibioticus
C>Date: 13-Jan-1995 #sequence_revision 06-Dec-1996 #text_change 16-Aug-2004
C/Accession: S43048; S41729
R:Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Submitted to the EMBL Data Library, February 1993
A:Reference number: S43048
A:Accession: S43048
A:Molecule type: DNA
A:Residues: 1-3519 <SWA>
A:Cross-references: UNIPROT:Q07017; EMBL:L09654; NID:G153407; PID:AAA19695.1; PID:G15340
R:Swan, D.G.; Rodriguez, A.M.; Vilches, C.; Mendez, C.; Salas, J.A.
Mol. Gen. Genet. 242, 358-362, 1994
A>Title: Characterisation of a Streptomyces antibioticus gene encoding a type I polyketic
A:Reference number: S41729; NUID:94150470; PMID:8107683
A:Accession: S41729
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1683-3238, 3273-3303, 'K', 3305-3407, 'T', 3409-3462, 'Y', 3464-3516, 'E', 3518-3519 <
A:Cross-references: EMBL:L09654
C:Genes:
C:Keyword: antibiotic biosynthesis; carrier protein
C:Keyword: antibiotic biosynthesis; carrier protein
F:59-462/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:570-851/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:1202-1181/Domain: short-chain alcohol dehydrogenase homology <SAD1>
F:1489-1560/Domain: acyl carrier protein homology <ACPI>
F:1708-2111/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:2221-2502/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:2858-3037/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F:3143-3214/Domain: acyl carrier protein homology <ACP2>
F:3305-3500/Domain: oleoyl-[acyl-carrier-protein] hydrolyase homology <ACPH>
Query Match 48.1%; Score 50; DB 2; Length 3519;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 3 VLVGVGSPNRRNEPLNG 18
DB 3401 VLVGVGSPNRRNEPLNG 3416
RESULT 2
C3GP Complement C3 precursor - guinea pig
N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit; (C3Spec): Cavia porcellus (guinea pig)
C>Date: 07-Feb-1992 #sequence_revision 07-Oct-1994 #text_change 09-Jul-2004
C/Accession: A3156; S03375; A20342; D20342; C20342; A31222

A:Residues: 1-1076 <NHA>
 A:Cross-references: EMBL:U41508; PIDN:AAA82623.1; CESP:W01C8.3
 C:Genetics:
 A:Gene: CESP:W01C8.3
 A:Introns: 59/3; 92/2; 157/3; 189/3; 220/2; 251/3; 275/2; 319/1; 374/3; 407/2

Query Match 44.2%; Score 46; DB 2; Length 1076;
 Best Local Similarity 63.6%; Pred. No. 45;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LVDCSPNRNE 14
 :|||:|
 DB 837 IIDGYHPRNE 847

RESULT 6

ubiquitinol-cytochrome-c reductase (EC 1.10.2.2) chain VIII [Imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 09-Jul-2004

C:Accession: T46746
 R:Labo:HaJdu G.; Braun, H.P.; Romp, N.; Grivell, L.A.; Berden, J.A.; Schmitz, U.K.
 Biochem. J. 320, 769-775, 1996
 A:Title: Subunit VII of ubiquitinol-cytochrome-c oxidoreductase from Neurospora crassa is
 A:Reference number: Z23142; MUID:97156980; PMID:9003361
 A:Accession: T46746

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-107 <LMB>

A:Cross-references: UNIPROT:P48503; EMBL:U20790; NID:9687738; PIDN:AAC49654.1; PID:96877

A:Experimental source: strain 74-OR23-1A; clone 5

C:Genetics:

A:Gene: OCR8
 C:Superfamily: Schistosacharomyces pombe ubiquitinol-cytochrome-c reductase chain VIII
 C:Keywords: electron transfer; membrane-associated complex; mitochondrion; oxidative pho

Query Match 43.3%; Score 45; DB 2; Length 107;
 Best Local Similarity 57.1%; Pred. No. 4.5;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 GYSPNRNEPTGNS 20
 :|||:|
 DB 41 GISPNRQNPICATY 54

RESULT 7
 JCS920
 potassium channel 2 - human
 C:Species: Homo sapiens (man)
 C:Date: 28-Mar-1998 #sequence_revision 28-Mar-1998 #text_change 09-Jul-2004

C:Accession: JCS920
 R:Su, K.; Kyaw, H.; Fan, P.; Zeng, Z.; Shell, B.K.; Carter, K.C.; Li, Y.

Biochem. Biophys. Res. Commun. 242, 675-681, 1997
 A:Title: Isolation, characterization, and mapping of two human potassium channels.

A:Reference number: JCS919
 A:Accession: JCS920

A:Molecule type: mRNA

A:Residues: 1-513 <SUA>
 A:Cross-references: UNIPROT:Q9UIK4; GB:AF033383; NID:92739502; PIDN:AAC05635.1; PID:9273

C:Comment: This plasma membrane protein has six transmembrane domains, and is involved in
 an external stimulus of neurotransmitters and neuropeptides.

C:Genetics:

A:Gene: KH2
 A:Map position: 20q13
 C:Superfamily: potassium channel protein drkl

Query Match 43.3%; Score 45; DB 2; Length 513;
 Best Local Similarity 47.4%; Pred. No. 28;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 SVLVDCSPNRNEPTGNS 20
 :|||:|
 DB 315 TLLVDGAAGRRKRGAGNS 333

RESULT 8

AE1021
 formate dehydrogenase (EC 1.2.1.2) - Salmonella enterica subsp. enterica serovar Typhi (

C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AE1021

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AE1021

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-715 <PAR>

A:Cross-references: UNIPROT:Q8Z1Q0; GB:AL513382; PIDN:CAD09270.1; PID:916505274; GSPDB:G

C:Keywords: oxidoreductase

Query Match 43.3%; Score 45; DB 2; Length 715;
 Best Local Similarity 47.4%; Pred. No. 41;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 SVLVDCSPNRNEPTGNS 20
 :|||:|
 DB 256 SKVBCYTBESVSEITGVS 274

RESULT 9
 XPB5A9
 large structural phosphoprotein p150 - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004

C:Accession: A29533; S09795
 R:Jahn, G.; Kouzarides, T.; Mach, M.; Scholl, B.C.; Plachter, B.; Traupe, B.; Preddie, E.

J. Virol. 61, 1358-1367, 1987
 A:Title: Map position and nucleotide sequence of the gene for the large structural phosph

A:Reference number: A29533; MUID:87198858; PMID:3033266

A:Accession: A29533

A:Molecule type: DNA

A:Residues: 1-1048 <JAH>

A:Cross-references: UNIPROT:P08318; GB:M16022; NID:9330643; PIDN:AAA45992.1; PID:9330644

A:Gene: M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.; Cerny, R.; Horzmann, T.; i

M.; Barrell, B.G.
 Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09795

A:Molecule type: DNA

A:Residues: 1-1048 <CHE>
 A:Cross-references: EMBL:X17403; NID:958591; PIDN:CAA35431.1; PID:959637

C:Genetics:
 A:Map position: 0.160-0.186
 C:Superfamily: human cytomegalovirus large structural phosphoprotein; large structural p

C:Keywords: phosphoprotein
 F;2-366/Domain: large structural phosphoprotein homology <CLS>

Query Match 43.3%; Score 45; DB 1; Length 1048;
 Best Local Similarity 70.0%; Pred. No. 63;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 DGSPNRNEP 15
 :|||:|
 DB 537 DGTPNRKDP 546


```

RESULT 10
T39379
sexual differentiation and meiosis protein ste20 - fission yeast (Schizosaccharomyces po
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39379; T46557
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, September 1995
A:Reference number: Z21849
A:Accession: T39379
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1309 <DEV>
A:Cross-references: UNIPROT:Q09743; EMBL:254140; PIDN:CAA90815.1; GSPDB:GN00066; SPDB:SF
A:Experimental source: strain 972h-, cosmid c12C2
R:Hiltl, N.; Baumann, D.; Edenhardter, E.; Stalder, M.; Schweingruber, M.E.
submitted to the EMBL Data Library, February 1998
A:Description: A new gene, ste20, in fission yeast Schizosaccharomyces pombe - necessary
and pH regulation.
A:Reference number: Z23064
A:Accession: T46557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1309 <HIL>
A:Cross-references: EMBL:AJ223984; PIDN:CAA11758.1
A:Gene: ste16; SPBC12C2.02c; ste20
A:Map position: 1
A:Function:
A:Description: necessary for sexual development and also responsible for sensitivity, me

Query Match
Best Local Similarity 43.3%; Score 45; DB 2; Length 1309;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 DGYSPNRNEPLTGN 20
:||||:|:|:|
Db 811 IDGYSEQVSEPIFSNS 826

RESULT 11
T16305
hypothetical protein F40F4.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16305
R:Wilson, R.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F40F4.
A:Reference number: Z18493
A:Accession: T16305
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2214 <WIL>
A:Cross-references: UNIPROT:Q20219; EMBL:U40420; NID:g1065513; PID:g1065514; PIDN:AAA814
C:Genetics:
A:Gene: CESP:F40F4.6
A:Introns: 57/3; 95/1; 302/3; 323/1; 380/1; 404/1; 468/1; 507/2; 547/3; 595/3; 1532/1; 1

Query Match
Best Local Similarity 43.3%; Score 45; DB 2; Length 2214;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 VLVDGSPNRNEPLNG 18
:|:|:|:|:|:|
Db 1950 VICNGRSPPTNPAIC 1965

RESULT 12
C81903
hypothetical protein NMA1343 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

```

```

C:Accession: C81903
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel]
; Holroyd, S.; Jagside, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: C81903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-186 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84591.1; PID:g738001;
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1343

Query Match
Best Local Similarity 42.3%; Score 44; DB 2; Length 186;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 6 DGYSPNRNEPLTGN 20
|||:|:|:|:|
Db 129 DGYTLPHHEAIAAGNA 143

RESULT 13
D8118
conserved hypothetical protein / ankyrin-related protein NMB1133, NMB1171 [imported] - Ne
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: D8118; AB1114
R:Retzlaff, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, B.K.; Hatt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
rt, H.; Qin, H.; Yamahevan, J.; Gyll, J.; Scarlato, V.; Maignan, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: D8118
A:Molecule type: DNA
A:Residues: 1-253 <TET>
A:Cross-references: UNIPROT:Q9URZ6; GB:AE002462; GB:AE002098; NID:g7226363; PIDN:AAF41523
A:Experimental source: serogroup B, strain MC58
A:Accession: AB1114
A:Molecule type: DNA
A:Residues: 1-253 <TE2>
A:Cross-references: GB:AE002465; GB:AE002098; NID:g7226401; PIDN:AAF41556.1; PID:g7226401
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1133; NMB1171

Query Match
Best Local Similarity 42.3%; Score 44; DB 2; Length 253;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 6 DGYSPNRNEPLTGN 20
|||:|:|:|:|
Db 196 DGYTLPHHEAIAAGNA 210

RESULT 14
AD2023
carboxyl-terminal proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AD2023
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yabate, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2023
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-445 <KUR>
A:Cross-references: UNIPROT:Q8YW78; GB:BA000019; PIDN:BA073437.1; PID:q17130828; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: a111738
C:Superfamily: carboxyl-terminal processing protease

Query Match 42.3%; Score 44; DB 2; Length 445;
Best Local Similarity 58.8%; Pred. No. 34;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SVLVDSYSPNRNEPLTG 18
:|||||:|||||
Db 312 AVLVDSNGSASASEILTG 328

RESULT 15

T35818
probable integral membrane export protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35818
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999
A:Reference number: Z21589
A:Accession: T35818
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-748 <MUR>
A:Cross-references: UNIPROT:Q92577; EMBL:AL035569; PIDN:CA837578.1; GSPDB:GN00070; SCOE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8D9.14

Query Match 42.3%; Score 44; DB 2; Length 748;
Best Local Similarity 53.3%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 VLVDSYSPNRNEPLT 17
:|||||:|||||
Db 408 LVLDSFGPGVNGEPLT 422

Search completed: October 27, 2005, 18:57:23
Job time : 13.3721 secs

(Order) 11/01/01 01/01/01

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:34:49 ; Search time 58.1395 Seconds
(without alignments)
176.155 Million cell updates/sec

Title: US-10-612-090-14
Perfect score: 104
Sequence: 1 SSVLDVGYSPNREPLTGENS 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	1148	2 Q9H7S7	Q9H7S7 homo sapien
2	104	100.0	6995	2 Q6GRK2	Q6GRK2 homo sapien
3	104	100.0	22152	2 Q8WK17	Q8WK17 homo sapien
4	55	52.9	258	2 Q9D1H1	Q9D1H1 mus musculi
5	50	48.1	3519	1 Q9L56	Q9L56 streptomyce
6	49	47.1	1666	1 CO3_CAVPO	P12387 cavia porce
7	48	46.2	1714	2 Q6I203	Q6I203 oestrinia nu
8	47	45.2	347	2 Q7VCG8	Q7VCG8 prochloroco
9	47	45.2	407	1 PPI_YEAST	P23250 saccharomyc
10	47	45.2	407	2 Q6Q5G6	Q6Q5G6 saccharomyc
11	47	45.2	867	2 Q9XBS2	Q9XBS2 zymomonas m
12	46	44.2	227	2 Q63WY6	Q63WY6 burkholderi
13	46	44.2	302	2 Q6CJ98	Q6CJ98 kluyveromyc
14	46	44.2	306	2 Q7NMG2	Q7NMG2 chromobacte
15	46	44.2	424	2 Q8AS18	Q8AS18 bacteroides
16	46	44.2	944	2 Q9GYG8	Q9GYG8 caenorhabdi
17	46	44.2	1018	1 M2R2_MOUSE	M2R2 mus musculi
18	46	44.2	1018	2 Q8BHS5	Q8BHS5 m homologue
19	46	44.2	1457	2 Q7R039	Q7R039 giardia lam
20	46	44.2	1723	2 Q95WK9	Q95WK9 lymantria d
21	45	43.3	107	1 UCRC_NEUCR	P48503 neurospora
22	45	43.3	167	1 HXD3_BRARE	O42310 brachydanio
23	45	43.3	242	2 Q89YJ4	Q89YJ4 bacteroides
24	45	43.3	271	2 Q8TV05	Q8TV05 methanopyru
25	45	43.3	324	2 Q7QH82	Q7QH82 anopheles g
26	45	43.3	325	2 Q7QH81	Q7QH81 anopheles g
27	45	43.3	449	2 Q9SR19	Q9SR19 arabidopsis
28	45	43.3	513	1 KCQ3_HUMAN	O9U144 homo sapien
29	45	43.3	513	2 Q86T85	Q86T85 homo sapien
30	45	43.3	715	2 Q821Q0	Q821Q0 salmonella
31	45	43.3	1042	2 Q7RE07	Q7RE07 plasmodium

32	45	43.3	1046	2 Q9DXH5	Q9DXH5 human cytom
33	45	43.3	1048	1 P100_HCMVA	P08318 human cytom
34	45	43.3	1048	2 Q7M6Q1	Q7M6Q1 human cytom
35	45	43.3	1049	2 Q6RX16	Q6RX16 human cytom
36	45	43.3	1049	2 Q6RW99	Q6RW99 human cytom
37	45	43.3	1309	1 ST16_SCHPO	Q09743 schizosacch
38	45	43.3	2214	2 Q20219	Q20219 caenorhabdi
39	44.5	42.8	586	2 Q884Z9	Q884Z9 pseudomonas
40	44	42.3	155	2 Q8RNB3	Q8RNB3 uncultured
41	44	42.3	200	2 Q853A6	Q853A6 mycobacteri
42	44	42.3	215	1 PCPI_RALSO	Q8X69 ralteconar s
43	44	42.3	253	1 YB33_NEIMB	Q9J26 neisseria m
44	44	42.3	253	1 YD43_NEIMA	Q9J22 neisseria m
45	44	42.3	259	2 Q17188	Q17188 bombyx mori

ALIGNMENTS

RESULT 1
ID Q9H7S7 PRELIMINARY; PRT; 1148 AA.
AC Q9H7S7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ14303.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta; DOI=10.1038/ng1285;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hoboliti T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimoriya K., Iehibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe S., Hirooka S., Chiba Y., Iehida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shichara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakabe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukumami Y.,
RA Fujimori Y., Komiyama M., Tashiro K., Tanigami A., Fujitawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okumoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK024365; BAB14899.1; -.
DR HSSP; Q9D1H1; IIVZ.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 7.
DR PROSITE; PS50024; SEA; 3.
SQ SEQUENCE 1148 AA; 127957 MW; 3861B0D5EFD8ABC CRC64;

```

Query Match          100.0%; Score 104; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 SSVLDGYSPPNREPLTGNS 20
Db 1070 SSVLDGYSPPNREPLTGNS 1089

RESULT 2
O96RK2 PRELIMINARY; PRT; 6995 AA.
AC O96RK2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE Mucin 16 (Fragment).
GN Name=MUC16;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347965; PubMed=11369781; DOI=10.1074/jbc.M103554200;
RA Yin B.W., Lloyd K.O.;
RT "Molecular cloning of the cal25 ovarian cancer antigen. Identification
as a new mucin, muc16."
RL J. Biol. Chem. 276:27371-27375(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX Lloyd K.O., Yin B.W.T.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF361486; AAK74120.3; -.
DR HSSB; Q9DIH1; 11VZ.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 20.
DR SMART; SM00200; SEA; 10.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00024; SEA; 6.
FT NON TER
SO SEQUENCE 6995 AA; 744958 MW; 80C797BDBF33A2B CRC64;

Query Match          100.0%; Score 104; DB 2; Length 6995;
Best Local Similarity 100.0%; Pred. No. 3.7e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 SSVLDGYSPPNREPLTGNS 20
Db 6917 SSVLDGYSPPNREPLTGNS 6936

RESULT 3
O8WX17 PRELIMINARY; PRT; 22152 AA.
AC O8WX17;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Ovarian cancer related tumor marker CA125.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21646939; PubMed=11786729; DOI=10.1159/000050638;
RA O'Brien T.J., Beard J.B., Underwood L.J., Dennis R.A., Santin A.D.,
RA York L.;
RT "The CA 125 gene: an extracellular superstructure dominated by repeat
sequences."
RL Tumour Biol. 22:348-366(2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA O'Brien T.J., Underwood L.J., Beard J.B.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144442; AAL65133.2; -.
DR Gene; HGNC:15582; MUC16.
DR Pfam; PF01390; SEA; 51.
DR SMART; SM00200; SEA; 23.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_2.
DR PROSITE; PS00024; SEA; 11.
SQ SEQUENCE 22152 AA; 235268 MW; B3E7BDF19997A440 CRC64;

Query Match          100.0%; Score 104; DB 2; Length 22152;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 SSVLDGYSPPNREPLTGNS 20
Db 22074 SSVLDGYSPPNREPLTGNS 22093

RESULT 4
O9DIH1 PRELIMINARY; PRT; 258 AA.
AC O9DIH1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched
library, clone:111000814 product:hypothetical SEA domain containing
protein, full insert sequence.
GN Name=111000814R1K;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN PANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RA The PANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Atawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama T., Nishi K., Kikunai T., Tashiro H., Itoh M.,

```

```

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.",
RL Genome Res. 10:1757-1771(2000).
[6]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=whole body;
RC Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carinchi P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hasegaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imorani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Komoto H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Taya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK003577; BAB2869.1; -.
DR PDB; 1IVZ; NMR; A=60-190.
DR MGI; MGI:1920982; 111008114R1K.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Hypothetical protein.
SQ SEQUENCE 258 AA; 29425 MW; B64D9B63394D84E7 CRC64;

Query Match 52.9%; Score 55; DB 2; Length 258;
Best Local Similarity 57.9%; Pred. No. 1.1;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 SVLVGYSPPRNPLETGN 20
||| ||| ||| |||
Db 181 SVFVDCYSONRDDVWKN 199

RESULT 5
OL56 STRAT STANDARD; PRT; 3519 AA.
AC 007017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
GN Name=orfB;
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150470; PubMed=8107683;
RA Swan D.G., Rodriguez A.M., Vilches C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type
RT I polyketide synthase which has an unusual coding sequence.",
Mol. Gen. Genet. 242:358-362(1994).
CC -1- FUNCTION: May be involved in the biosynthesis of the oleandomycin
CC lactone ring.
CC -1- COFACTOR: Contains 2 covalently bound phosphopantetheines.
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; L09654; AAA19695.1; -.
DR PIR; S43048; S43048.
DR HSSP; Q03133; 1MO2.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR009081; ACP_like.
DR InterPro; IPR000794; ketoacyl_synth.
DR InterPro; IPR006163; Pp bind.
DR InterPro; IPR006162; Pantoate S.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00698; Acyl_transf_1; 2.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt C; 2.
DR Pfam; PF00550; Pp-binding; 2.
DR Pfam; PF00975; Thioesterase; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
KW Acyltransferase; Antibiotic biosynthesis; Multifunctional enzyme;
KW NADP. Phosphopantetheine; Repeat; Transferase.
SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAAB61F86 CRC64;

Query Match 48.1%; Score 50; DB 1; Length 3519;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 VLVGYSPPRNPLETGN 18
||| ||| ||| |||
Db 3401 VLVGVPPGROBPVPG 3416

RESULT 6
COS_CAVPO STANDARD; PRT; 1666 AA.
ID COS_CAVPO
AC P12387;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Complement C3 precursor [Contains: C3a anaphylatoxin].
GN Name=C3;
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathii; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90307998; PubMed=1973176;
RA Auebach H.S., Burger R., Dods A., Colten H.R.;
RT "Molecular basis of complement C3 deficiency in guinea pigs.",
RL J. Clin. Invest. 86:96-106(1990).
CC [2]
CC SEQUENCE OF 676-753.
CC MEDLINE=89113342; PubMed=3064079;
CC Gerard N.P., Lively M.O., Gerard C.;

```

RT "Amino acid sequence of guinea pig C3a anaphylatoxin.";
 RL Protein Seq. Data Anal. 1:473-478(1988).
 RN [3]
 RP SEQUENCE OF 993-1032.
 RX MEDLINE=83178889; PubMed=6838833;
 RA Thomas M.L., Tack B.F.;
 RT "Identification and alignment of a thiol ester site in the third component of guinea pig complement.";
 RL Biochemistry 22:942-947(1983).
 CC -1- FUNCTION: C3 plays a central role in the activation of the complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently, via its reactive thioester, to cell surface carbohydrates or immune aggregates.
 CC -1- FUNCTION: Derived from proteolytic degradation of complement C3, C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes.
 CC -1- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg residues, forming two chains, beta and alpha, linked by a disulfide bond. C3 convertase activates C3 by cleaving the alpha chain, releasing C3a anaphylatoxin and generating C3b (beta chain + alpha' chain).
 CC -1- SIMILARITY: Contains 1 anaphylatoxin-like domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; M34054; AAA37038.1; -
 DR PIR; A37156; C3GP.
 DR HSSP; P01026; 10QF.
 DR InterPro; IPR002890; A2M_N.
 DR InterPro; IPR009048; AM_receptor_bind.
 DR InterPro; IPR000020; Anaphylatoxin.
 DR InterPro; IPR001840; Anaphylatoxin.
 DR InterPro; IPR008964; Invasin_intimin.
 DR InterPro; IPR001599; MacroglobinA2.
 DR InterPro; IPR001134; Netrin_C.
 DR InterPro; IPR008930; Texp_Cyc_toroid.
 DR InterPro; IPR008993; TIMP_like.
 DR Pfam; PF00207; A2M; 1.
 DR Pfam; PF01835; A2M_N; 1.
 DR Pfam; PF01821; ANATO; 1.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PR00004; ANAPHYLATOXN.
 DR PRODOM; PD003264; Anaphylatoxin; 1.
 DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
 DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
 DR PROSITE; PS50189; NTR; 1.
 DR Complement alternate pathway; Complement pathway;
 KW Direct protein sequencing; Glycoprotein; Inflammatory response;
 KW Plasma; Signal; Thioester bond.
 FT SIGNAL 1 29
 FT CHAIN 30 1666 Complement C3.
 FT CHAIN 30 671 Complement C3 beta chain.
 FT CHAIN 676 1666 Complement C3 alpha chain.
 FT PEPTIDE 676 753 C3a anaphylatoxin.
 FT CHAIN 754 1666 Complement C3b alpha' chain.
 FT DOMAIN 688 723 Anaphylatoxin-like.
 FT DOMAIN 1522 1664 NTR.
 FT SITE 753 754 Cleavage (by C3 convertase).
 FT DISULFID 557 821 Interchain (By similarity).
 FT DISULFID 630 666 By similarity.
 FT DISULFID 698 725 By similarity.
 FT DISULFID 699 732 By similarity.

FT DISULFID 712 733 By similarity.
 FT DISULFID 878 1517 By similarity.
 FT DISULFID 1106 1163 By similarity.
 FT DISULFID 1363 1493 By similarity.
 FT DISULFID 1394 1462 By similarity.
 FT DISULFID 1510 1515 By similarity.
 FT DISULFID 1522 1593 By similarity.
 FT DISULFID 1540 1649 By similarity.
 FT DISULFID 1640 1648 By similarity.
 FT CROSSLINK 1015 1018 Isoglutamyl cysteine thioester (Cys-Gln).
 FT CARBOHYD 944 944 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 1620 1620 D->N (in Ref. 2).
 FT CONFLICT 731 731 Missing (in Ref. 3).
 FT CONFLICT 1013 1013 Q->E (in Ref. 2).
 FT CONFLICT 1018 1018 Missing (in Ref. 3).
 FT CONFLICT 1031 1031 Missing (in Ref. 3).
 SQ SEQUENCE 1666 AA; 186487 MW; 1C1F1219944AFD49 CRC64;
 Query Match 47.1%; Score 49; DB 1; Length 1666;
 Best Local Similarity 55.6%; Pred. No. 1.1e+02;
 Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 Oy 3 VLVDGSPNRPNEPLTNGS 20
 Db 305 VLVDGSPNRPNEPLTNGS 322
 RESULT 7
 ID 061203 PRELIMINARY; PRT; 1714 AA.
 AC 061203;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE Cadherin A1.
 OS Oestrinia nubilalis (European corn borer).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyralidae;
 OC Pyralidae; Pyraustinae; Oestrinae.
 OX NCBI_TaxId=29057;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Coates B.S., Sumnerford D.V., Hellmich R.L.;
 RL Submitted (Apr-2004) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: Contains 11 cadherin domains.
 DR EMBL; AY612336; AAT37678.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR002126; Cadherin.
 DR Pfam; PF00028; Cadherin; 3.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 10.
 DR PROSITE; PS00232; CADHERIN_1; 2.
 DR PROSITE; PS50268; CADHERIN_2; 11.
 KW Calcium; Calcium-binding.
 SQ SEQUENCE 1714 AA; 191900 MW; 89E3CAE0614654DE CRC64;
 Query Match 46.2%; Score 48; DB 2; Length 1714;
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Oy 3 VLVDGSPNRPNEPLTNGS 20
 Db 1014 VLPEIYADRDDEPDTNS 1031
 RESULT 8
 ID 07VCG8 PRELIMINARY; PRT; 347 AA.
 AC 07VCG8;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TREMblrel. 26, last annotation update)
 DE Predicted ATPase.
 GN OrderedLocNames=Pro0772;
 OS Prochlorococcus marinus.
 OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
 CC Prochlorococcus.
 NCBI_TaxID=1219;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SARG / CCM 1375 / SS120;
 RX MEDLINE=22810154; PubMed=12917466; DOI=10.1073/pnas.1733211100;
 RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,
 Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,
 Makarova K.S., Ostrowski M., Ozta S., Robert C., Rogozin I.B.,
 Scanlan D.J., Tandeau de Marsac N., Weissbach J., Wincker P.,
 Wolf Y.I., Hesse W.R.;
 RT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,
 a nearly minimal oxyphototrophic genome";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).
 DR EMBL; AB017163; AAP9816.1; -.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR001307; Thiosulfatrans.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS00380; RHODANSE_1; UNKNOWN_1.
 DR PROSITE; PS0206; RHODANSE_3; 1.
 DR Complete proteome.
 SQ SEQUENCE 347 AA; 39187 MW; 388F2B1A2BE4956 CRC64;
 QY
 DB 258 LVDSYSPHKNEDL 270
 Query Match 45.2%; Score 47; DB 2; Length 347;
 Best Local Similarity 69.2%; Pred. No. 35;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 9
 RPI1_YEAST STANDARD; PRT; 407 AA.
 ID P23250;
 AC 01-NOV-1991 (Rel. 20, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Negative RAS protein regulator protein.
 GN Name=RPI1; OrderedLocNames=YIL119C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91304380; PubMed=1649384;
 RA Kim J.-H., Powers S.;
 RT "Overexpression of RPI1, a novel inhibitor of the yeast Ras-cyclic AMP
 pathway, down-regulates normal but not mutationally activated ras
 function.";
 RT Mol. Cell. Biol. 11:3894-3904(1991).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=6288C / AB972;
 RX MEDLINE=97313266; PubMed=9169870;
 RA Churche C.M., Bowman S., Badcock K., Bankier A., Brown D.,
 Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
 Moutre S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
 RL Nature 387:84-87(1997).
 CC -!- FUNCTION: Negative regulator of the Ras-cyclic AMP pathway.
 CC Negatively regulate the activity of normal but not mutationally
 CC activated Ras proteins. The down-regulatory effect of RPI1
 CC requires the presence of one of the two Ras GTPase activators,

CC IRA1 and IRA2.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 1 Myb-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 246833; CAA6873.1; -
 CC EMBL; M63178; AAA5001.1; -
 CC EMBL; S42682; AAA39843.1; -
 CC PIR; S49890; S49890.
 CC GeneMOnline; 139654; -
 CC SGD; S000001381; RPI1.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:0005083; F:small GTPase regulatory/interacting protein. .; IMP.
 CC GO; GO:0007265; P:RAS protein signal transduction; IMP.
 CC InterPro; IPR001005; Myb DNA binding.
 CC PROSITE; PS00037; MYB_1; FALSE_NEG.
 CC PROSITE; PS00334; MYB_2; FALSE_NEG.
 CC PROSITE; PS00090; MYB_3; 1.
 CC DNA-binding; Nuclear Protein.
 FT DNA BIND 90 158
 FT MYB.
 FT N-> K (in Ref. 1).
 FT 37 37
 FT CONFLICT 71 73
 FT SNS-> P (in Ref. 1).
 FT 91 91
 FT CONFLICT 170 174
 FT A-> V (in Ref. 1).
 FT 174 174
 FT CONFLICT 240 240
 FT Missing (in Ref. 1).
 FT 240 240
 FT CONFLICT 242 242
 FT C-> S (in Ref. 1).
 FT 242 242
 FT CONFLICT 252 252
 FT N-> S (in Ref. 1).
 FT 252 252
 FT CONFLICT 252 252
 FT N-> NNNSNN (in Ref. 1).
 FT 252 252
 SQ SEQUENCE 407 AA; 46623 MW; 6C6A18DCD04F58701 CRC64;
 QY
 DB 26 YSPNLTPTTNS 37
 Query Match 45.2%; Score 47; DB 1; Length 407;
 Best Local Similarity 66.7%; Pred. No. 43;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 10
 YSPNLTPTTNS PRELIMINARY; PRT; 407 AA.
 ID Q605G6;
 AC 0605G6;
 DT 05-JUL-2004 (TREMblrel. 27, Created)
 DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMblrel. 27, Last annotation update)
 DE YIL119C.
 GN Name=RPI1;
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Marischky G., Rolfs A., Richardson A., Kane M., Bagui M., Taycher E.,
 Hu Y., Vanhegel F., Weger J., Kramer J., Moreira D., Kelley F.,
 Zuo D., Raphael J., Hogle C., Jepson D., Williamson J., Camargo A.,
 Gonzalez L., Vasconcelos A.T., Simpson A., Kolodner R., Harlow E.,
 Labber J.;
 RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AY558088; AAS56414.1; -
 DR SGD; S000001381; RPI1.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005677; F:DNA binding; IEA.
 DR InterPro; IPR001005; Myb DNA binding.
 DR PROSITE; PS00090; MYB_3; 1.
 SQ SEQUENCE 407 AA; 46633 MW; 448DC302815B6FD CRC64;

```

Query Match          45.2%; Score 47; DB 2; Length 407;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 8 YSPNREPLTGN 19
DB 26 YSPNINPTITSN 37

RESULT 11
OQXBS2
ID OQXBS2; PRELIMINARY; PRT; 867 AA.
AC OQXBS2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Membrane alanyl aminopeptidase.
GN Name=pepN;
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Um H.W., Kang H.S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF157493; AAD42403.1; -.
DR MEROPS; M01.005; -.
DR GO; GO:0004177; F:aminopeptidase activity; IEA.
DR GO; GO:0004179; F:membrane alanyl aminopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001930; Peptidase_M1.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR PRINTS; PR00756; ALADIPITASE.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR Aminopeptidase.
SQ SEQUENCE 867 AA; 97500 MW; BDD94757AC90ECC CRC64;

Query Match          45.2%; Score 47; DB 2; Length 867;
Best Local Similarity 60.0%; Pred. No. 1e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 VLVDGYSPPNRNPLT 17
DB 713 VLVDGYSPPNRKALT 727

RESULT 12
OQ3WY6
ID OQ3WY6; PRELIMINARY; PRT; 227 AA.
AC OQ3WY6;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BPSL0753;
OS Burkholderia pseudomallei K96243.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=272560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K96243;
RX PubMed=15377794;
RA Holden M.T.G., Tiltball R.W., Peacock S.J., Cerdano-Tarraga A.M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
RA Bentley S.D., Sebaihia M., Thomson N.R., Baason N., Beacham I.R.,
RA Brooks K., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosset B., Davis P., Deshazer D.,
RA Felwell T., Fraser A., Hance Z., Hauser C., Holtroyd S., Jørgels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,

```

```

RA Rabinowitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songstivalai S., Stevens K., Tumapa S., Vesaratchavee M.,
RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
DR EMBL; BX571965; CAH34746.1; -.
KW Hypothetical protein.
SQ SEQUENCE 227 AA; 24840 MW; D98FA4F04FECB059 CRC64;

Query Match          44.2%; Score 46; DB 2; Length 227;
Best Local Similarity 61.5%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 VDGYSPNRNPLT 17
DB 121 IDGYSKNSEPNT 133

RESULT 13
OQCU98
ID OQCU98; PRELIMINARY; PRT; 302 AA.
AC OQCU98;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Similar to egd|S0005690 Saccharomyces cerevisiae YOR164C.
GN ORFNames=KLAD0F202959;
OS Kluyveromyces lactic NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boitame A., Boyer J., Catolico L., Confantolero F., de Darvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennou D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bojotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Galliardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382126; CAG98699.1; -.
DR InterPro; IPR007317; DUF410.
DR Pfam; PF04190; DUF410; 1.
SQ SEQUENCE 302 AA; 34920 MW; 0C295541A6165E00 CRC64;

Query Match          44.2%; Score 46; DB 2; Length 302;
Best Local Similarity 56.2%; Pred. No. 44;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 VLVDGYSPPNRNPLT 18
DB 95 VLVDGQEPNLDVYTG 110

RESULT 14
QTNMG2

```

```

ID 07NMG2 PRELIMINARY; PRT; 306 AA.
AC 07NMG2;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Chemotaxis motB protein.
GN Name=mocB2; OrderedLocustNames=CV2027;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=536;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12472 / DSM 30191;
RC MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimarães C.T.,
Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
Alves-Gomes J.A., Andrade E.M., Aratipe J., de Araujo M.F.F.,
Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
Baptista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burily H.A.,
Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Chuelre L.M.O.,
Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
Fantinetti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
Gazzinelli R.T., Gomes E.A., Gonçalves P.R., Grangeiro T.B.,
Gratcapaglia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,
Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,
di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
Santos E.B.P., Santos F.R., Schneider M.P.C., Seunarez H.N.,
Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Umenyi T.,
Vettore A., Waesem R., Zaha A., Simpson A.J.G.;
" The complete genome sequence of Chromobacterium violaceum reveals
remarkable and exploitable bacterial adaptability." ;
Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
EMBL: AE016917; AA059659.1; -
GO: GO:0009779; C:outer membrane (sensu Gram-negative Bacteria); IEA.
DR InterPro: IPR006665; OmpA/MocB.
DR Pfam: PF00691; OmpA; 1.
DR ProDom: PD000930; OmpA/MocB; 1.
KM Complete proteome.
SQ SEQUENCE 306 AA; 33083 MW; D2F4EDBDF832A18 CRC64;

```

```

Query Match 44.2%; Score 46; DB 2; Length 306;
Best Local Similarity 44.4%; Pred. No. 45;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 SSVLVGYSPPNREPLTG 18
DB 273 SSVLVGYSPPNREPLTG 290

```

```

RESULT 15
08A5L8
ID 08A5L8 PRELIMINARY; PRT; 424 AA.
AC 08A5L8;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=BT2220;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VP1-5482 / ATCC 29148;
RC MEDLINE=22550858; PubMed=12563928; DOI=10.1126/science.1080029;
XU J., Bjurvell M.K., Hmrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis." ;
Science 299:2074-2076(2003).
EMBL: AE016935; AA077327.1; -
DR EMBL: AE016935; AA077327.1; -
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 424 AA; 47549 MW; 3C87947B0184AC3E CRC64;

```

```

Query Match 44.2%; Score 46; DB 2; Length 424;
Best Local Similarity 40.0%; Pred. No. 66;
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 SSVLVGYSPPNREPLTGS 20
DB 358 SSVLVGYSPPNREPLTGS 377

```

Search completed: October 27, 2005, 18:55:07
Job time : 63.1395 secs

The Page Book (1999)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:31:33 ; Search time 69.2093 Seconds
(without alignments)
111.765 Million cell updates/sec

Title: US-10-612-090-15
Perfect score: 105
Sequence: 1 TNYQRKRNIEDALNQLFRN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	ADJ66742	Adj66742 Human Muc
2	105	100.0	108	ADJ66729	Adj66729 Human Muc
3	105	100.0	108	ADJ66740	Adj66740 Human Muc
4	105	100.0	178	ABP30979	Abp30979 Predicted
5	105	100.0	178	ADA08642	Ada08642 Human O77
6	105	100.0	178	Adf08985	Adf08985 Secreted
7	105	100.0	233	ABP30978	Abp30978 Extracellular
8	105	100.0	233	ADA08641	Ada08641 Human O77
9	105	100.0	233	ADP08984	Adp08984 Secreted
10	105	100.0	284	ABU54859	Abu54859 Human CA1
11	105	100.0	318	ABP31025	Abp31025 Amino acil
12	105	100.0	318	ADA08747	Ada08747 Human O77
13	105	100.0	318	Adf09090	Adf09090 Secreted
14	105	100.0	367	AAE12631	Aae12631 Human gen
15	105	100.0	438	AAE12555	Aae12555 Human ova
16	105	100.0	438	AAE12623	Aae12623 Human gen
17	105	100.0	438	ABP30973	Abp30973 Partial p
18	105	100.0	438	ABP30899	Abp30899 0772P c10
19	105	100.0	438	ADA08636	Ada08636 Human O77
20	105	100.0	438	ADA08543	Ada08543 Human ova
21	105	100.0	438	Adf08886	Adf08886 Secreted
22	105	100.0	438	Adf08979	Adf08979 Secreted
23	105	100.0	438	ADG46173	Adg46173 Human ova
24	105	100.0	439	ABU54709	Abu54709 Human CA1
25	105	100.0	526	AAE12634	Aae12634 Human gen

26	105	100.0	748	8	ADS94302	Ads94302 CA 125/O7
27	105	100.0	809	3	ADS94303	Ads94303 CA 125/O7
28	105	100.0	833	3	AAE12554	Aae12554 Human ova
29	105	100.0	833	5	ABP30898	Abp30898 0772P c10
30	105	100.0	833	7	ADA08542	Ada08542 Human ova
31	105	100.0	833	7	Adf08885	Adf08885 Secreted
32	105	100.0	833	7	ADG46172	Adg46172 Human ova
33	105	100.0	914	3	AAE12552	Aae12552 Human ova
34	105	100.0	914	4	ABP99203	Abp99203 Human ova
35	105	100.0	914	5	ABP30968	Abp30968 Hypothetel
36	105	100.0	914	5	ABP30896	Abp30896 0772P pro
37	105	100.0	914	7	ADA08631	Ada08631 Human O77
38	105	100.0	914	7	ADA08465	Ada08465 Human ova
39	105	100.0	914	7	ADP08974	Adp08974 Secreted
40	105	100.0	914	7	ADP08808	Adp08808 Secreted
41	105	100.0	914	7	ADG46095	Adg46095 Human ova
42	105	100.0	914	8	ADN40451	Adn40451 Human bre
43	105	100.0	1148	4	AAE12583	Aae12583 Human pro
44	105	100.0	1148	4	ABP50283	Abp50283 HOST-1 ov
45	105	100.0	1148	5	ABG96380	Abg96380 Human ova

ALIGNMENTS

RESULT 1
ADJ66742
ID ADJ66742 standard; peptide; 20 AA.
XX
XX ADJ66742;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human Muc1/Muc16 protein-related peptide SegID15.
XX
XX monoclonal antibody; epitope; non-shed extracellular portion;
XX shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
XX breast cancer; ovarian cancer.
XX
XX Homo sapiens.
XX
XX WO2004005470-A2.
XX
XX PD
XX 15-JAN-2004.
XX
XX PF 03-JUL-2003; 2003WO-US020907.
XX
XX PR 03-JUL-2002; 2002US-0393094P.
XX
XX (IMMU-) IMMUNOGEN INC.
XX
XX Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
XX Vater CA;
XX
XX WPI; 2004-091350/09.
XX
XX PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
XX treating or monitoring malignancies, such as breast or ovarian cancer.
XX
XX PS Claim 14; SEQ ID NO 15; 113pp; English.
XX
XX This invention relates to a novel isolated monoclonal antibody that
XX specifically binds to an epitope of a non-shed extracellular portion of a
XX shed antigen or of human Muc1 or Muc16 protein. The invention may be
XX useful for the development of compounds with a cytostatic or for gene
XX therapy. The composition and methods are useful in diagnosing, treating
XX or monitoring malignancies, such as breast or ovarian cancer. The present
XX sequence is that of a peptide which was used for raising antibodies to
XX the extracellular, non-shed region of Muc16 and which is claimed in the
XX specification.
XX
XX Sequence 20 AA;
XX
XX SQ

Query Match 100.0%; Score 105; DB 8; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.6e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TNYORNRKNIEDALNQLFRN 20
 |||
 1 TNYORNRKNIEDALNQLFRN 20

RESULT 2

ADJ66729
 ID ADJ66729 standard; protein; 108 AA.

AC ADJ66729;

DT 06-MAY-2004 (first entry)

DE Human Muc16 epitope amino acid sequence.

KM monoclonal antibody; epitope; non-shed extracellular portion;

KW shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;

KM breast cancer; ovarian cancer.

OS Homo sapiens.

PN WO2004005470-A2.

PD 15-JAN-2004.

PF 03-JUL-2003; 2003WO-US020907.

PR 03-JUL-2002; 2002US-0393094P.

PA (IMMU-) IMMUNOGEN INC.

PI Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,

PI Vater CA;

DR WPI; 2004-091350/09.

PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,

PT treating or monitoring malignancies, such as breast or ovarian cancer.

PS Claim 13; SEQ ID NO 2; 113pp; English.

CC This invention relates to a novel isolated monoclonal antibody that

CC specifically binds to an epitope of a non-shed extracellular portion of a

CC shed antigen or of human Muc1 or Muc16 protein. The invention may be

CC useful for the development of compounds with a cytostatic or for gene

CC therapy. The composition and methods are useful in diagnosing, treating

CC or monitoring malignancies, such as breast or ovarian cancer. The present

CC sequence is that of a Muc epitope of the invention.

XX Sequence 108 AA;

XX Query Match 100.0%; Score 105; DB 8; Length 108;

XX Best Local Similarity 100.0%; Pred. No. 6.3e-09;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORNRKNIEDALNQLFRN 20

DB 1 TNYORNRKNIEDALNQLFRN 20

RESULT 3

ADJ66740
 ID ADJ66740 standard; protein; 108 AA.

AC ADJ66740;

DT 06-MAY-2004 (first entry)

DE Human Muc16 GST fusion protein amino acid sequence SeqID13.

XX monoclonal antibody; epitope; non-shed extracellular portion;

KW shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;

KW breast cancer; ovarian cancer.

OS Homo sapiens.

PN WO2004005470-A2.

PD 15-JAN-2004.

PF 03-JUL-2003; 2003WO-US020907.

PR 03-JUL-2002; 2002US-0393094P.

PA (IMMU-) IMMUNOGEN INC.

PI Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,

PI Vater CA;

DR WPI; 2004-091350/09.

PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,

PT treating or monitoring malignancies, such as breast or ovarian cancer.

PS Disclosure; SEQ ID NO 13; 113pp; English.

CC This invention relates to a novel isolated monoclonal antibody that

CC specifically binds to an epitope of a non-shed extracellular portion of a

CC shed antigen or of human Muc1 or Muc16 protein. The invention may be

CC useful for the development of compounds with a cytostatic or for gene

CC therapy. The composition and methods are useful in diagnosing, treating

CC or monitoring malignancies, such as breast or ovarian cancer. The present

CC sequence is that of the human Muc16 GST fusion protein which was used in

CC the exemplification of the invention.

XX Sequence 108 AA;

XX Query Match 100.0%; Score 105; DB 8; Length 108;

XX Best Local Similarity 100.0%; Pred. No. 6.3e-09;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORNRKNIEDALNQLFRN 20

DB 1 TNYORNRKNIEDALNQLFRN 20

RESULT 4

ABP30979

ID ABP30979 standard; protein; 178 AA.

AC ABP30979;

DT 02-JUL-2002 (first entry)

DE Predicted extracellular domain of 0772P.

KW Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.

OS Homo sapiens.

PN WO200206317-A2.

PD 24-JAN-2002.

PF 17-JUL-2001; 2001WO-US022635.

PR 17-JUL-2000; 2000US-00617747.

PR 10-AUG-2000; 2000US-00636801.

PR 20-SEP-2000; 2000US-00667857.

PR 04-APR-2001; 2001US-00827271.

PR 18-JUN-2001; 2001US-00884441.

PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX WPI; 2002-164781/21.
DR
XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
XX Example 13; Page 361-362; 408pp; English.
XX
CC This invention relates to polypeptides comprising an immunogenic portion
CC of an ovarian carcinoma protein which acts as an immunostimulant and is
CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations, and antigen presenting cells that express
CC the polypeptides are useful for stimulating an immune response in a
CC patient and treating ovarian cancer. This sequence represents protein
CC related to the invention
XX
SQ Sequence 178 AA:
Query Match 100.0%; Score 105; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TNYQRKNRNIEDALNQLFRN 20
DB 70 TNYQRKNRNIEDALNQLFRN 89
RESULT 5
ADA08642
ID ADA08642 standard; protein; 178 AA.
XX
XX ADA08642;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human O772P partial protein #12.
XX
XX Human O772P partial protein #12.
XX
XX human; gene therapy; ovarian cancer; cancer.
XX
OS Homo sapiens.
XX
XX US2003091580-A1.
XX
XX 15-MAY-2003.
XX
XX 17-JUL-2001; 2001US-00907969.
XX
XX 18-JUN-2001; 2001US-00884441.
XX
XX (MITC/) MITCHAM J L.
XX (KING/) KING G E.
XX (ALGA/) ALGATE P A.
XX (FLIN/) FLING S P.
XX (RETT/) RETTER M W.
XX (FANG/) FANGER G R.
XX (REED/) REED S G.
XX (VEDV/) VEDVICK T S.
XX (CART/) CARTER D.
XX (HILL/) HILL P.
XX (ALBO/) ALBONE E.
XX
XX Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
XX Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2003-532352/50.
XX
XX New isolated O772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian

PT cancer.
XX
XX Disclosure; Page 28; 371pp; English.
XX
XX The invention relates to an isolated O772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma cDNAs and protein cDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma
CC antigen.
XX
SQ Sequence 178 AA:
Query Match 100.0%; Score 105; DB 7; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TNYQRKNRNIEDALNQLFRN 20
DB 70 TNYQRKNRNIEDALNQLFRN 89
RESULT 6
ADF08985
ID ADF08985 standard; protein; 178 AA.
XX
XX ADF08985;
XX
DT 12-FEB-2004 (first entry)
XX
DE Secreted ovarian carcinoma antigen seqid 489.
XX
XX gene therapy; protein therapy; vaccine; antibody inhibition;
XX breast cancer; restorative therapy; diagnostic agent; immunoassay;
XX secreted ovarian carcinoma antigen.
XX
XX Homo sapiens.
XX
XX US2003124140-A1.
XX
XX 03-JUL-2003.
XX
XX 17-JUL-2002; 2002US-00198053.
XX
XX 17-DEC-1998; 98US-00215681.
XX 17-DEC-1998; 98US-00216003.
XX 23-JUN-1999; 99US-00338933.
XX 24-SEP-1999; 99US-00404879.
XX 17-JUL-2000; 2000US-00617747.
XX 10-AUG-2000; 2000US-00636801.
XX 20-SEP-2000; 2000US-00667857.
XX 04-APR-2001; 2001US-00827271.
XX 18-JUN-2001; 2001US-00884441.
XX 17-JUL-2001; 2001US-00907969.
XX
XX (CORI-) CORIXA CORP.
XX
XX Bangur CS, Retter MW, Fanger GR, Hill P;
XX WPI; 2003-697152/82.
XX
XX Oncogenic nucleic acids useful for the prevention, diagnosis and
PT treatment of breast cancer.
XX
XX Example 13; SEQ ID NO 489; 399pp; English.
XX
XX The invention describes nucleic acids (I) and the polypeptides (II) they
CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
CC treating diseases related to their aberrant expression i.e. breast
CC cancers. For example, (I) and (II) may be used to treat disorders

CC	associated with decreased expression by rectifying mutations or deletions
CC	in a patient's genome that affect the activity of (II) by expressing
CC	inactive proteins or to supplement the patients own production of (II).
CC	Additionally, (I) may be used to produce (II), by inserting (I) into a
CC	host cell and culturing the cell to express the protein (II). (I) And its
CC	complementary sequences may also be used as DNA probes in diagnostic
CC	assays to detect and quantitate the presence of similar nucleic acids in
CC	samples, and therefore which patients may be in need of restorative
CC	therapy. The host cell may also be used as antigens in the production of
CC	antibodies against (II) and in assays to identify modulators of (II)'s
CC	expression and activity. The anti-(II) antibodies, agonists and
CC	antagonists may be used to regulate expression and activity and as
CC	diagnostic agents for detecting the presence of (II) in samples (e.g. by
CC	immunoassay). This sequence represents a secreted ovarian carcinoma
CC	antigen.
XX	
SQ	Sequence 178 AA:
Query Match	100.0%; Score 105; DB 7; Length 178;
Best Local Similarity	100.0%; Pred. No. 1,1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0,	
OY	1 TNNQRKNIEDALNQLFRN 20
Db	70 TNYQRKRNIEDALNQLFRN 89
RESULT 7	
ID	ABP30978
AC	ABP30978 standard; protein; 233 AA.
XX	
AC	ABP30978;
XX	
DT	02-JUL-2002 (first entry)
XX	
DE	Extracellular and transmembranal regions of 0772P.
XX	
KM	Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
OS	Homo sapiens.
XX	
PN	WO200206317-A2.
PD	24-JAN-2002.
XX	
PX	17-JUL-2001; 2001WO-US022635.
XX	
PR	17-JUL-2000; 2000US-00617747.
PR	10-AUG-2000; 2000US-00636801.
PR	20-SEP-2000; 2000US-00667857.
PR	04-APR-2001; 2001US-00827271.
PR	18-JUN-2001; 2001US-00884441.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI	Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
DR	WPI; 2002-164781/21.
XX	
PT	Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT	protein or its variants, useful for stimulating an immune response in a
XX	patient and treating ovarian cancer.
PS	Example 13; Page 361; 408pp; English.
XX	
CC	This invention relates to polypeptides comprising an immunogenic portion
CC	of an ovarian carcinoma protein which acts as an immunostimulant and is
CC	cyclostatic. The polypeptides, polynucleotides, antibodies, fusion
CC	proteins, T cell populations and antigen presenting cells that express
CC	the polypeptides are useful for stimulating an immune response in a
CC	patient and treating ovarian cancer. This sequence represents protein
CC	related to the invention

```

XX      SQ      Sequence 233 AA;
Query Match          100.0%; Score 105; DB 5; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.ee-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY           1 TNYORNRKNIEDALNQLFRN 20
              |||ooooooooooooo|||
Db            70 TNYORNRKNIEDALNQLFRN 89

RESULT 8
ADA08641
ID ADA08641 standard; protein; 233 AA.
AC ADA08641;
DT 06-NOV-2003 (first entry)
DE Human O772P partial protein #11.
KM human; gene therapy; ovarian cancer; cancer.
OS Homo sapiens.
PN US2003091580-A1.
PD 15-MAY-2003.
PF 17-JUL-2001; 2001US-00907969.
PR 18-JUN-2001; 2001US-00884441.
PA (MITC/) MITCHAM J L.
PA (KING/) KING G E.
PA (ALGA/) ALGATE P A.
PA (FLIN/) FLING S P.
PA (RETT/) RETTER M W.
PA (FANG/) FANGER G R.
PA (REED/) REED S G.
PA (VEDV/) VEDVICK T S.
PA (CAR/) CARTER D.
PA (HILL/) HILL P.
PA (ALBO/) ALBONE E.
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
DR WP1: 2003-532352/50.
PT New isolated O772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g., ovarian
PT cancer.
PS Example 13; Page 28; 371pp; English.
CC The invention relates to an isolated O772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma cDNAs and protein cDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma
CC antigen.
CC CC
SQ      Sequence 233 AA;
Query Match          100.0%; Score 105; DB 7; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


QY 1 TNYQRKRNIEDALNQLFRN 20
 |||||
 DB 70 TNYQRKRNIEDALNQLFRN 89

RESULT 9

ID ADF08984 standard; peptide; 233 AA.

AC ADF08984;

DT 12-FEB-2004 (first entry)

DE Secreted ovarian carcinoma antigen seqid 488.

KW gene therapy; protein therapy; vaccine; antibody inhibition;
 KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
 KW secreted ovarian carcinoma antigen.

OS Homo sapiens.

PN US2003124140-A1.

PD 03-JUL-2003.

PF 17-JUL-2002; 2002US-00198053.

PR 17-DEC-1998; 98US-00215681.

PR 17-DEC-1998; 98US-00216003.

PR 23-JUN-1999; 99US-00338933.

PR 24-SEP-1999; 99US-00404879.

PR 17-JUL-2000; 2000US-00617747.

PR 10-AUG-2000; 2000US-00636801.

PR 20-SEP-2000; 2000US-00657857.

PR 04-APR-2001; 2001US-00827271.

PR 18-JUN-2001; 2001US-00884441.

PR 17-JUL-2001; 2001US-00907969.

PA (CORI-) CORIXA CORP.

PI Bangur CS, Retter MW, Fanger GR, Hill P;

DR WPI; 2003-897152/82.

PT Oncogenic nucleic acids useful for the prevention, diagnosis and

PT treatment of breast cancer.

PS Example 13; SEQ ID NO 488; 399pp; English.

CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acid (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancers. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patient's own production of (II).
 CC Additionally, (II) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) And its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigens in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and
 CC antagonists may be used to regulate expression and activity and as
 CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunoassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.

XX Sequence 233 AA;

Query Match 100.0%; Score 105; DB 7; Length 233;
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TNYQRKRNIEDALNQLFRN 20
 |||||
 DB 70 TNYQRKRNIEDALNQLFRN 89

RESULT 10

ID ABUS4859 standard; protein; 284 AA.

AC ABUS4859;

DT 12-MAR-2003 (first entry)

DE Human CA125 carboxy terminal domain.

KW Human; CA125; protein repeat; chromosome 19q 13.2; amino terminal domain;
 KW amino terminal extension; carboxy terminal domain; vaccine; cancer;
 KW ovarian cancer; carcinoma.

OS Homo sapiens.

PN WO200283866-A2.

PD 24-OCT-2002.

PF 12-APR-2002; 2002WO-US011734.

PR 17-APR-2001; 2001US-0284175P.

PR 19-JUN-2001; 2001US-0299380P.

PR 27-SEP-2001; 2001US-00965738.

PR 21-DEC-2001; 2001US-0345180P.

PA (UYAR-) UNIV ARKANSAS.

PI O'brien T, Beard J, Underwood L;

DR WPI; 2003-093013/08.

PT New CA125 molecules, useful as a gold standard for detecting and

PT monitoring the presence of CA125 antigen which can be used for

PT diagnosing, monitoring or treating patients with cancer or for developing

PT vaccine against cancer.

PS Claim 1; Fig 9b; 694pp; English.

CC The invention relates to a CA125 protein comprising: (a) an extracellular
 CC amino terminal domain; (b) an amino terminal extension; (c) a multiple
 CC repeat domain; and (d) a carboxy terminal domain. The extracellular amino
 CC terminal domain comprises 5 genomic exons, the amino terminal extension
 CC comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and
 CC the carboxy terminal domain comprises a transmembrane anchor with a short
 CC cytoplasmic domain, and further comprises 9 genomic exons. The gene for
 CC CA125 is located on human chromosome 19q 13.2. Also included are isolated
 CC CA125 repeat domains, nucleic acids (including variants, homologues and
 CC degenerate versions) encoding CA125 proteins or repeat units, a vector
 CC comprising the nucleic acid, a cultured cell comprising the vector, a
 CC method of expressing CA125 antigen in a cell, the amino acid sequences of
 CC the CA125 repeat units (or their variants, fragments or sequences 50%
 CC identical to them), a purified antibody that selectively binds to an
 CC epitope in the receptor-binding domain of CA125 protein, a diagnostic for
 CC detecting and monitoring the presence of CA125 antigen (comprising
 CC recombinant CA125 having at least one repeat unit of the CA125 repeat
 CC domain including epitope binding sites), a therapeutic vaccine to treat
 CC mammals with elevated CA125 antigen levels or at risk of developing a
 CC disease or disease recurrence associated with elevated CA125 antigen
 CC levels (comprising recombinant CA125 repeat domains including epitope
 CC binding sites) and an antisense oligonucleotide that inhibits the
 CC expression of CA125. The CA125 molecule, particularly the multiple repeat
 CC domains are useful as a gold standard for detecting and monitoring the
 CC presence of CA125 antigen, which can be used for diagnosing, monitoring
 CC or treating patients with ovarian cancer and other carcinomas where CA125

CC is expressed. The molecules are also useful for developing a vaccine
 CC against cancer. The present sequence is a CA125 repeat protein
 XX
 SQ Sequence 284 AA;

Query Match 100.0%; Score 105; DB 6; Length 284;
 Best Local Similarity 100.0%; Pred. No. 2e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TNYQRNKNIEDALNQLFRN 20
 |||||
 DB 121 TNYQRNKNIEDALNQLFRN 140

RESULT 11
 ABP31025
 ID ABP31025 standard; protein; 318 AA.
 XX
 AC ABP31025;

DT 02-JUL-2002 (first entry)

DE Amino acid sequence of the 3' constant region of 0772P.

KW Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.

XX Homo sapiens.

OS WO200206317-A2.

PN 24-JAN-2002.

PF 17-JUL-2001; 2001WO-US022635.

PR 17-JUL-2000; 2000US-00617747.

PR 10-AUG-2000; 2000US-00636801.

PR 20-SEP-2000; 2000US-00657857.

PR 04-APR-2001; 2001US-00827271.

PR 18-JUN-2001; 2001US-00884441.

XX (CORI-) CORIXA CORP.

PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;

PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;

DR WPI; 2002-164781/21.

PS Claim 3; Page 399-400; 408pp; English.

CC This invention relates to polypeptides comprising an immunogenic portion

CC of an ovarian carcinoma protein which acts as an immunostimulant and is

CC cytotoxic. The polypeptides, polynucleotides, antibodies, fusion

CC proteins, T cell populations and antigen presenting cells that express

CC the polypeptides are useful for stimulating an immune response in a

CC patient and treating ovarian cancer. This sequence represents protein

CC related to the invention

XX Sequence 318 AA;

Query Match 100.0%; Score 105; DB 5; Length 318;
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TNYQRNKNIEDALNQLFRN 20
 |||||
 DB 155 TNYQRNKNIEDALNQLFRN 174

RESULT 12

ADA08747
 ID ADA08747 standard; protein; 318 AA.
 XX
 AC ADA08747;

DT 06-NOV-2003 (first entry)
 DE Human 0772P constant region consensus sequence #1.
 XX
 KW human; gene therapy; ovarian cancer; cancer.

OS Homo sapiens.

PN US2003091580-A1.

PD 15-MAY-2003.

PF 17-JUL-2001; 2001US-00907969.

PR 18-JUN-2001; 2001US-00884441.

XX (MINC/) MITCHAM J L.

PA (KING/) KING G E.

PA (ALGA/) ALGATE P A.

PA (FLIN/) FLING S P.

PA (RETT/) RETTER M W.

PA (FANG/) FANGER G R.

PA (REED/) REED S G.

PA (VEDV/) VEDVICK T S.

PA (CART/) CARTER D.

PA (HILL/) HILL P.

PA (ALBO/) ALBONE E.

XX Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;

PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;

DR WPI; 2003-532352/50.

XX New isolated 0772P polypeptides and polynucleotides, useful in gene

PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian

PT cancer.

XX Claim 3; SEQ ID NO 594; 371pp; English.

PS The invention relates to an isolated 0772P polypeptide, which has the

CC structure fully defined in the specification. The composition containing

CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells

CC or antigen presenting cells are useful for stimulating an immune response

CC and treating ovarian cancer. Detecting the presence of the

CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian

CC carcinoma cDNAs and protein cDNAs were identified using microarray

CC technology. The present sequence represents a human ovarian carcinoma

CC antigen.

XX Sequence 318 AA;

Query Match 100.0%; Score 105; DB 7; Length 318;
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TNYQRNKNIEDALNQLFRN 20
 |||||
 DB 155 TNYQRNKNIEDALNQLFRN 174

RESULT 13

ADFO9090

ID ADFO9090 standard; protein; 318 AA.

XX ADFO9090;

DT 12-FEB-2004 (first entry)

03-JAN-2002 (first entry)
Human gene 1 encoded secreted protein fragment, SEQ ID NO: 27.
Human; ovarian cancer antigen; proliferative disorder; cancer; tumour;
respiratory system disorder; asthma; haematopoietic disorder; skin aging;
immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
rheumatoid arthritis; inflammation; neurological disorder; septic shock;
Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
epithelial cell proliferation; transplantation; chemotaxis; infection;
food additive; wound healing; endocrine disorder; kidney disorder;
gene therapy; cytostatic.
Homo sapiens.
MO200170804-A1.
27-SEP-2001.
16-MAR-2001; 2001WO-US008585.
17-MAR-2000; 2000US-0190076P.
23-AUG-2000; 2000US-0227009P.
(HUMA-) HUMAN GENOME SCI INC.
Birre CE, Rosen CA;
WPI, 2001-639119/73.
Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
metastases.
Disclosure; Page 9; 427pp; English.
AAD20618-AAAD20624 represent cDNAs corresponding to 7 human ovarian and
ovarian cancer associated protein (collectively known as ovarian cancer
antigens) genes, and AAE12623-MAE12629 represent the proteins they
encode. AAE12630-AAE12638 represent human ovarian cancer antigenic
fragments or variants. Ovarian cancer antigens and their corresponding
cDNAs are used in the prevention, diagnosis and treatment of diseases
associated with their inappropriate expression. These disorders include
proliferative disorders, cancer, tumours, respiratory system disorders,
asthma, haematopoietic disorders, diseases of the immune system, AIDS,
skin disorders, autoimmune diseases (e.g., rheumatoid arthritis),
inflammation, allergies, neurological disorders (e.g., Alzheimer's
disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
cardiovascular disorders, angiogenic disorders, kidney disorders,
gastrointestinal disorders, endocrine disorders and infections. The
c proteins can also be used to aid wound healing and epithelial cell
proliferation, to prevent skin aging due to sunburn, to maintain organs
before transplantation, to identify their cognate ligands or binding
partners, in chemotaxis and can be used as a food additive. Antibodies
specific for a protein of the invention can be used in alleviating
symptoms associated with the disorders mentioned above and in diagnostic
immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
the invention is used in gene therapy. The present sequence represents a
human ovarian cancer antigenic fragment of the invention

Sequence 367 AA:

Query Match 100.0%; Score 105; DB 4; Length 367;
Best Local Similarity 100.0%; Pred. No. 2..7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TNYQNRKNTEDALNQLFRN 20
|||||
db 275 TNYQRNRKNTEDALNQLFRN 294

```

RESULT 15
AAB12555
ID AAB12555 standard; protein; 438 AA.
XX
XX AAB12555;
AC
DT 07-NOV-2000 (first entry)
XX
XX 07-NOV-2000 (first entry)
DE Human ovarian carcinoma antigen O772P protein SEQ ID NO:390.
XX
XX Human; ovarian carcinoma; ovarian cancer; therapy; diagnosis;
KM tumour antigen; identification; cytostatic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
XX MO20036107-A2.
XX
XX 22-JUN-2000.
XX
XX 17-DEC-1999; 99WO-US030270.
XX
XX 17-DEC-1998; 98US-00215681.
XX
XX 17-DEC-1998; 98US-00216003.
XX
XX 23-JUN-1999; 99US-00338933.
XX
XX 24-SEP-1999; 99US-00404879.
XX
XX (CORI-) CORIXA CORP.
XX
XX Mitcham JL, King GE, Algate PA, Frudakis TN;
XX
XX WPI; 2000-431589/37.
XX
XX N-PSDB; AAA70076.
XX
XX Immunogenic portion of an ovarian carcinoma protein and the nucleic acid
XX encoding it useful for the diagnosis, prevention and treatment of
XX cancer, preferably ovarian cancer.
XX
XX Example 2; Page 203-204; 299pp; English.
XX
XX The present invention describes an isolated polypeptide comprising an
XX immunogenic portion of an ovarian carcinoma protein (or its variants).
XX Ovarian carcinoma proteins, and polynucleotides encoding them, have
XX cytostatic activity and can be used in gene therapy and vaccines. Ovarian
XX carcinoma polypeptides, nucleic acids, antibodies and vaccines are useful
XX for the prevention, diagnosis and treatment of cancer, preferably ovarian
XX cancer. AA69691 to AA70077 and AAB12552 to AAB12557 represent human
XX ovarian carcinoma polynucleotides and proteins used in the
XX exemplification of the present invention
XX
XX SQ Sequence 438 AA;
XX
XX Query Match 100.0%; Score 105; DB 3; Length 438;
XX Best Local Similarity 100.0%; Pred. No. 3.3e-08;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TNYQRKRNIEDALNQLFRN 20
XX ||||||||||||||||
XX 275 TNYQRKRNIEDALNQLFRN 294

```

Search completed: October 27, 2005, 18:44:24
 Job time : 70.2093 secs

Sequence 7616, Ap

NUMBER OF SEC T

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-667-857-390

Query Match          100.0%; Score 105; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNKRNIEDALNQLFRN 20
DB 275 TNYORKNKRNIEDALNQLFRN 294

RESULT 3
US-09-404-879A-389
; Sequence 389, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon B.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-389

Query Match          100.0%; Score 105; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNKRNIEDALNQLFRN 20
DB 670 TNYORKNKRNIEDALNQLFRN 689

RESULT 4
US-09-667-857-389
; Sequence 389, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon B.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-667-857-389
```

```
Query Match          100.0%; Score 105; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNKRNIEDALNQLFRN 20
DB 670 TNYORKNKRNIEDALNQLFRN 689

RESULT 5
US-09-404-879A-312
; Sequence 312, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon B.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-312

Query Match          100.0%; Score 105; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNKRNIEDALNQLFRN 20
DB 751 TNYORKNKRNIEDALNQLFRN 770

RESULT 6
US-09-338-933-312
; Sequence 312, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-338-933-312

Query Match          100.0%; Score 105; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNKRNIEDALNQLFRN 20
DB 751 TNYORKNKRNIEDALNQLFRN 770

RESULT 7
US-09-667-857-312
; Sequence 312, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
```

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Rector, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriek
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-667-857-312

```

```

Query Match      100.0%; Score 105; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 TNYQRKNRNEALNQLFRN 20
Db      751 TNYQRKNRNEALNQLFRN 770

```

```

RESULT 8
US-09-404-879A-388
; Sequence 388, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 388
; LENGTH: 772
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-404-879A-388

```

```

Query Match      65.7%; Score 69; DB 4; Length 772;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 TNYQRKNRNEALNQLFRN 13
Db      751 TNYQRKNRNEALNQLFRN 763

```

```

RESULT 9
US-09-667-857-388
; Sequence 388, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Rector, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.

```

```

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriek
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 388
; LENGTH: 772
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-667-857-388

```

```

Query Match      65.7%; Score 69; DB 4; Length 772;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 TNYQRKNRNEALNQLFRN 13
Db      751 TNYQRKNRNEALNQLFRN 763

```

```

RESULT 10
US-09-270-767-42105
; Sequence 42105, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42105
; LENGTH: 516
; TYPE: PRF
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-42105

```

```

Query Match      43.8%; Score 46; DB 4; Length 516;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 TNYQRKNRNEALNQLFRN 20
Db      132 SNYRRIKQIMDILKMKFS 151

```

```

RESULT 11
US-09-248-796A-17941
; Sequence 17941, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17941
; LENGTH: 336
; TYPE: PRF
; ORGANISM: Candida albicans

```

US-09-248-796A-17941

Query Match 43.3%; Score 45.5; DB 4; Length 336;

Best Local Similarity 64.7%; Pred. No. 22;

Matches 11; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 2 YORNRKNIEDALNQLF 18

Db 62 NYNNKRNRI---LNQWF 75

RESULT 12

US-09-115-150-5

; Sequence 5, Application US/09115150

; Patent No. 6503747

; GENERAL INFORMATION:

; APPLICANT: Kathariou, Sophia

; TITLE OF INVENTION: Serotype-Specific Probes for *Listeria Monocytogenes*

; FILE REFERENCE: A65378/RPT/DBS

; CURRENT APPLICATION NUMBER: US/09/115,150

; CURRENT FILING DATE: 1998-07-14

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 316

; TYPE: PRT

; ORGANISM: Bacterial

US-09-115-150-5

Query Match 42.9%; Score 45; DB 4; Length 316;

Best Local Similarity 50.0%; Pred. No. 25;

Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 1 TNYNRKNIEDALNQLF 20

Db 121 TYTENKRNFDGNSLVN 140

RESULT 13

US-09-949-016-6984

; Sequence 6984, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: C1001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6984

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-6984

Query Match 42.4%; Score 44.5; DB 4; Length 596;

Best Local Similarity 64.7%; Pred. No. 61;

Matches 11; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 YORNRKNIEDALNQLF 19

Db 554 YORNR-LEKALQOLAR 569

RESULT 14

US-09-949-016-7843

; Sequence 7843, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: C1001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6984

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-6984

Query Match 42.4%; Score 44.5; DB 4; Length 596;

Best Local Similarity 64.7%; Pred. No. 61;

Matches 11; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 YORNRKNIEDALNQLF 19

Db 554 YORNR-LEKALQOLAR 569

RESULT 15

US-09-489-039A-7928

; Sequence 7928, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *KLEBSIELLA*

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7928

; LENGTH: 334

; TYPE: PRT

; ORGANISM: *Klebsiella pneumoniae*

US-09-489-039A-7928

Query Match 41.9%; Score 44; DB 4; Length 334;

Best Local Similarity 53.3%; Pred. No. 38;

Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 YORNRKNIEDALNQL 17

Db 116 YRKNMSINDELSQL 130

Search completed: October 27, 2005, 17:34:32

Job time : 18.3953 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:30:08 ; Search time 64 Seconds
(without alignments)
130.638 Million cell updates/sec

Title: US-10-612-090-15
Perfect score: 105
Sequence: 1 TNYORKNRNIEDALNQLFRN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1865214 seqs, 418043040 residues

Total number of hits satisfying chosen parameters: 1865214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US10F_PUBCOMB.pep.*
19: /cgn2_6/ptodata/2/pubppa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	US-10-612-090-15	Sequence 15, Appl
2	105	100.0	108	US-10-612-090-2	Sequence 2, Appl
3	105	100.0	108	US-10-612-090-13	Sequence 13, Appl
4	105	100.0	178	US-09-884-441-489	Sequence 489, App
5	105	100.0	178	US-09-907-969-489	Sequence 489, App
6	105	100.0	178	US-10-198-053-489	Sequence 489, App
7	105	100.0	178	US-10-860-790-489	Sequence 489, App
8	105	100.0	233	US-09-884-441-488	Sequence 488, App
9	105	100.0	233	US-09-907-969-488	Sequence 488, App
10	105	100.0	233	US-10-198-053-488	Sequence 488, App
11	105	100.0	233	US-10-860-790-488	Sequence 488, App

12	105	100.0	284	10	US-09-965-738-300	Sequence 300, App
13	105	100.0	318	10	US-09-907-969-594	Sequence 594, App
14	105	100.0	318	14	US-10-198-053-594	Sequence 594, App
15	105	100.0	318	17	US-10-860-790-594	Sequence 594, App
16	105	100.0	367	17	US-10-333-900-27	Sequence 27, Appl
17	105	100.0	367	18	US-10-858-412-225	Sequence 225, App
18	105	100.0	438	9	US-09-884-441-483	Sequence 483, App
19	105	100.0	438	9	US-09-884-441-483	Sequence 483, App
20	105	100.0	438	10	US-09-907-969-390	Sequence 390, App
21	105	100.0	438	10	US-09-907-969-483	Sequence 483, App
22	105	100.0	438	10	US-09-827-271-380	Sequence 380, App
23	105	100.0	438	14	US-10-198-053-390	Sequence 390, App
24	105	100.0	438	14	US-10-333-900-19	Sequence 19, Appl
25	105	100.0	438	15	US-10-333-900-30	Sequence 30, Appl
26	105	100.0	438	17	US-10-860-790-390	Sequence 390, App
27	105	100.0	438	17	US-10-860-790-483	Sequence 483, App
28	105	100.0	439	10	US-09-965-738-148	Sequence 148, App
29	105	100.0	456	18	US-10-858-412-226	Sequence 226, App
30	105	100.0	526	15	US-10-333-900-30	Sequence 30, Appl
31	105	100.0	545	14	US-10-243-243A-4	Sequence 4, Appl
32	105	100.0	583	14	US-10-142-515-4	Sequence 4, Appl
33	105	100.0	748	17	US-10-687-035-1	Sequence 1, Appl
34	105	100.0	809	17	US-10-687-035-2	Sequence 2, Appl
35	105	100.0	833	9	US-09-884-441-389	Sequence 389, App
36	105	100.0	833	10	US-09-907-969-389	Sequence 389, App
37	105	100.0	833	10	US-09-827-271-389	Sequence 389, App
38	105	100.0	833	14	US-10-198-053-389	Sequence 389, App
39	105	100.0	833	17	US-10-860-790-389	Sequence 389, App
40	105	100.0	914	9	US-09-778-320-206	Sequence 206, App
41	105	100.0	914	9	US-09-910-689-206	Sequence 206, App
42	105	100.0	914	9	US-09-884-441-312	Sequence 312, App
43	105	100.0	914	9	US-09-884-441-478	Sequence 478, App
44	105	100.0	914	10	US-09-907-969-112	Sequence 112, App
45	105	100.0	914	10	US-09-907-969-478	Sequence 478, App

ALIGNMENTS

RESULT 1
US-10-612-090-15
; Sequence 15, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 15
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-612-090-15

Query Match 100.0%; Score 105; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYORKNRNIEDALNQLFRN 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 TNYORKNRNIEDALNQLFRN 20

RESULT 2
US-10-612-090-2
; Sequence 2, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:

```

1  APPLICANT: Immunogen, Inc.
2  TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
3  FILE REFERENCE: A8340
4  CURRENT APPLICATION NUMBER: US/10/612,090
5  CURRENT FILING DATE: 2003-07-03
6  PRIOR APPLICATION NUMBER: US 60/393,094
7  PRIOR FILING DATE: 2002-07-03
8  NUMBER OF SEQ ID NOS: 33
9  SOFTWARE: PatentIn version 3.2
10 SEQ ID NO 2
11     LENGTH: 108
12     TYPE: PRT
13 ORGANISM: Homo sapiens
14 US-10-612-090-2

```

Query Match	100.0%;	Score 105;	DB 15;	Length 108;
Best Local Similarity	100.0%;	Pred. No. 7e-08;		
Matches 20; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy 1 TNYQRNKRNIEDALNQLFRN 20
1 |||||
Db 1 TNYQRNKRNIEDALNQLFRN 20

RESULT 3
US-10-612-090-13
; Sequence 13, Application US/10612090
; Publication No. US20040057952A1
General Information

```

? TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREO
? FILE REFERENCE: A8340
? CURRENT APPLICATION NUMBER: US/10/612,090
? PRIOR FILING DATE: 2003-07-03
? PRIOR APPLICATION NUMBER: US 60/393,094
? PRIOR FILING DATE: 2002-07-03
? NUMBER OF SEQ ID NOS: 33
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 13
? LENGTH: 108
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Fusion protein
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (1)..(1)
? OTHER INFORMATION: Glutathione S-transferase fusion site
? US-10-612-090-13

```

Query Match	100.0%;	Score 105;	DB 15;	Length 108;
Best Local Similarity	100.0%;	Pred. No. 7e-08;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	TNYQRNKRNI	EDALNQLFRN	20
Db	1	TNYQRNKRNI	EDALNQLFRN	20

```

RESULT 4
US-09-884-441-489
: Sequence 489, Application US/09884441
: Patent No. US20020119158A1
: GENERAL INFORMATION:
: APPLICANT: Algate, Paul A.
: APPLICANT: Carter, Patrick
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
: FILE REFERENCE: 210121,46CCT
: CURRENT APPLICATION NUMBER: US/09/884,441
: CURRENT FILING DATE: 2001-06-18
: NUMBER OF SEQ ID NOS: 489
: SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-489

```

```

Query Match      100.0%; Score 105; DB 9; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY      1 TNYQRNKNIEDALNQLFRN 20
          |||||
Db      70 TNYQRNKNIEDALNQLFRN 89
```

RESULT 5

; Sequence 489, Application US/09907969
; Publication No. US20030091580A1

```

1  APPLICANT: Mitcham, Jennifer L.
2  APPLICANT: King, Gordon E.
3  APPLICANT: Algate, Paul A.
4  APPLICANT: Rling, Steven P.
5  APPLICANT: Reiter, Marc W.
6  APPLICANT: Fanger, Gary Richard
7  APPLICANT: Reed, Steven G.
8  APPLICANT: Vedavick, Thomas S.
9  APPLICANT: Carter, Darrick
10 APPLICANT: Hill, Paul
11 APPLICANT: Albone, Earl
12 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
13 TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
14 FILE REFERENCE: 210121.46268
15 CURRENT APPLICATION NUMBER: US/09/907,969
16 CURRENT FILING DATE: 2001-07-17
17 NUMBER OF SEQ ID NOS: 596
18 SOFTWARE: FASTSEQ FOR Windows Version 4.0
19 SEQ ID NO 489
20 LENGTH: 178
21 TYPE: PRT
22 ORGANISM: Homo sapiens
23 US-09-907-969-489

```

Query Match	100.0%	Score 105;	DB 10	Length 178;
Best Local Similarity	100.0%	Pred. No. 1	2e-07;	
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      1 TNYQRNKNIEDALNQLFRN 20
        |||||
Db      70 TNYQRNKNIEDALNQLFRN 89
```

RESULT 6
US-10-19

```

? Sequence 489, Application US/10198053
? Publication No. US2003012440A1
? GENERAL INFORMATION:
? APPLICANT: Bangur, Chaitanya S.
? APPLICANT: Retter, Marc W.
? APPLICANT: Fanger, Gary R.
? APPLICANT: Hill, Paul
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
? OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
? FILE REFERENCE: 210121.46209
? CURRENT APPLICATION NUMBER: US/10/198,053
? CURRENT FILING DATE: 2002-07-17
? NUMBER OF SEQ. ID NOS: 624
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 489
? LENGTH: 178
? TYPE: PRT
? ORGANISM: Homo sapiens

```

US-10-198-053-489

Query Match 100.0%; Score 105; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRKNRIEDALNQLFRN 20
DB 70 TNYQRKNRIEDALNQLFRN 89

RESULT 7

US-10-860-790-489
; Sequence 489, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-860-790-489

Query Match 100.0%; Score 105; DB 17; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRKNRIEDALNQLFRN 20
DB 70 TNYQRKNRIEDALNQLFRN 89

RESULT 8

US-09-884-441-488
; Sequence 488, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; FILE REFERENCE: 210121.462C7
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-884-441-488

Query Match 100.0%; Score 105; DB 9; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRKNRIEDALNQLFRN 20
DB 70 TNYQRKNRIEDALNQLFRN 89

RESULT 9

US-09-907-969-488

; Sequence 488, Application US/09907969

; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-907-969-488

Query Match 100.0%; Score 105; DB 10; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRKNRIEDALNQLFRN 20
DB 70 TNYQRKNRIEDALNQLFRN 89

RESULT 10

US-10-198-053-488
; Sequence 488, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-198-053-488

Query Match 100.0%; Score 105; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRKNRIEDALNQLFRN 20
DB 70 TNYQRKNRIEDALNQLFRN 89

RESULT 11

US-10-860-790-488
; Sequence 488, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-488

Query Match          100.0%; Score 105; DB 17; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRNKNIEDALNQLFRN 20
DB 70 TNYQRNKNIEDALNQLFRN 89

RESULT 12
US-09-965-738-300
; Sequence 300, Application US/09965738
; Publication No. US2003014367A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy
; TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic a
; TITLE OF INVENTION: Therapeutic Interventions
; FILE REFERENCE: 40715-258841
; CURRENT APPLICATION NUMBER: US/09/965,738
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/284,175
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 300
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-738-300

Query Match          100.0%; Score 105; DB 10; Length 284;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRNKNIEDALNQLFRN 20
DB 121 TNYQRNKNIEDALNQLFRN 140

RESULT 13
US-09-907-969-594
; Sequence 594, Application US/0907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Pling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darwick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

```

```

; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 594
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-09-907-969-594

Query Match          100.0%; Score 105; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRNKNIEDALNQLFRN 20
DB 155 TNYQRNKNIEDALNQLFRN 174

RESULT 14
US-10-198-053-594
; Sequence 594, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 594
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-10-198-053-594

Query Match          100.0%; Score 105; DB 14; Length 318;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TNYQRNKNIEDALNQLFRN 20
DB 155 TNYQRNKNIEDALNQLFRN 174

RESULT 15
US-10-860-790-594
; Sequence 594, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02

```


(648) 742-0211

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 18:13:49 ; Search time 12.3721 Seconds
(without alignments)
155.538 Million cell updates/sec

Title: US-10-612-090-15

Perfect score: 105

Sequence: 1 TNYQRNKNIEDALNQLFRN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir79:*
2: pir1:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	48.6	841	2	T36384
2	48	45.7	666	2	T17396
3	47	44.8	1955	1	AGCH
4	46	43.8	535	2	S74703
5	46	43.8	1328	2	T43060
6	45	42.9	728	2	D84647
7	44	41.9	56	2	G81826
8	44	41.9	420	2	B72386
9	44	41.9	680	2	G82526
10	44	41.9	705	2	F70352
11	44	41.9	809	2	E90016
12	44	41.9	971	2	D70128
13	44	41.9	1879	2	T19481
14	43.5	41.4	475	2	H64218
15	43.5	41.4	2206	2	G71611
16	43	41.0	96	2	H70340
17	43	41.0	181	2	I40146
18	43	41.0	257	2	B70246
19	43	41.0	288	2	T35695
20	43	41.0	445	2	C71636
21	43	41.0	476	2	A28439
22	43	41.0	1240	2	T30834
23	42.5	40.5	94	2	H69887
24	42.5	40.5	469	2	T34173
25	42	40.0	188	2	A70256
26	42	40.0	201	2	E97049
27	42	40.0	278	2	D87517
28	42	40.0	297	2	C70251
29	42	40.0	302	2	S50609

30	42	40.0	306	2	S61169	hypothetical prote
31	42	40.0	334	2	D86892	hypothetical prote
32	42	40.0	343	2	C55590	hypothetical prote
33	42	40.0	384	2	H89873	hypothetical prote
34	42	40.0	430	2	UC1230	DNA-binding prote
35	42	40.0	510	2	AG1539	oligo-1,6-glucosid
36	42	40.0	571	2	S58356	pept protein - Sta
37	42	40.0	605	2	S64410	probable membrane
38	42	40.0	680	2	C82905	hypothetical prote
39	42	40.0	906	2	B96901	uncharacterized co
40	42	40.0	1214	2	T21915	hypothetical prote
41	42	40.0	1279	2	E64709	type IIS restricti
42	42	40.0	2269	2	T28677	thoxy protein -
43	41	39.0	108	2	T27649	hypothetical prote
44	41	39.0	225	2	T44720	hypothetical prote
45	41	39.0	232	2	H87113	conserved hypothec

ALIGNMENTS

RESULT 1
T36384
probable ATP-binding proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T36384
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: 221573
A:Accession: T36384
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-841 <OLI>
A:Cross-references: UNIPROT:Q9S6T8; EMBL:AL049628; PIDN:CAB40873.1; GSPDB:GNO0070; SCODE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCB94.24C
C:Superfamily: endopeptidase Clp ATP-binding chain
C:Keywords: ATP; molecular chaperone; nucleotide binding; P-loop
F:212-219/Region: nucleotide-binding motif A (P-loop)
F:279-284/Region: nucleotide-binding motif B
F:549-556/Region: nucleotide-binding motif A (P-loop)
F:617-622/Region: nucleotide-binding motif B
F:218/Binding site: ATP (Lys) #status predicted
F:555/Binding site: ATP (Lys) #status predicted

Query Match 48.6%; Score 51; DB 2; Length 841;
Best local similarity 47.4%; Pred. No. 13;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 TNYQRNKNIEDALNQLFR 19
DB 685 SNYERKNKVDDELKQHR 703

RESULT 2
T17396
vlp protein - Dichelobacter nodosus
C:Species: Dichelobacter nodosus
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17396
R:Billington, S.J.; Huggins, A.S.; Johansen, P.A.; Crellin, P.K.; Cheung, J.K.; Katz, M.
Infect. Immun. 67, 1277-1286, 1999
A>Title: Complete nucleotide sequence of the 27-kilobase virulence related locus (vrl) of
A:Reference number: Z18734; MUID:99150261; PMID:10024571
A:Accession: T17396
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-666 <Bl>
A:Cross-references: UNIPROT:Q32495; EMBL:U20246; NID:g349323; PID:g2317809; PIDN:AAC333
A:Experimental source: strain A198
C:Superfamily: Dichelobacter nodosus vlp protein

Query Match 45.7% Score 48; DB 2; Length 666;
 Best Local Similarity 55.6%; Pred. No. 29;
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 TNYQRNKNEDALNQLF 18
 |||:|||||:
 Db 345 TBYQTRREIRGDAIINTIF 362

RESULT 3
 AGCH
 agrin precursor - chicken
 C|Species: Gallus gallus (chicken)
 C|Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
 C|Accession: JH0591; A38857; B38857; I50692
 R|Tslm, K.W.K.; Ruegg, M.A.; Escher, G.; Kroege, S.; McMahon, U.J.
 Neuron 8, 677-689, 1992
 A|Title: cDNA that encodes active agrin.
 A|Reference number: JH0591; MUID:92232297; PMID:1314620
 A|Accession: JH0591
 A|Molecule type: mRNA
 A|Residues: 1-1955 <TS1>
 A|Cross-references: UNIPROT:P31696; GB:M94271; NID:g211120; PIDN:AAA48565.1; PID:g211121
 A|Experimental source: brain
 R|Ruegg, M.A.; Tslm, K.W.K.; Horton, S.E.; Kroege, S.; Escher, G.; Gensch, E.M.; McMahon
 Neuron 8, 691-699, 1992
 A|Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
 A|Reference number: A38857; MUID:92232298; PMID:1314621
 A|Contents: alternative splicing
 A|Accession: A38857
 A|Molecule type: mRNA
 A|Residues: 1132-1783; 1795-1955 <R2>
 A|Cross-references: GB:M97371
 A|Accession: B38857
 A|Molecule type: mRNA
 A|Residues: 1221-1647; 1653-1783; 1794-1955 <R3>
 A|Cross-references: GB:M97372
 A|Note: translation of the nucleotide sequence is not complete
 R|Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
 Dev. Biol. 158, 523-535, 1993
 A|Title: Developmental expression and alternative splicing of chick agrin RNA.
 A|Reference number: I50692; MUID:93345745; PMID:8393816
 A|Accession: I50692
 A|Status: preliminary; translated from GB/EMBL/DBJ
 A|Molecule type: mRNA
 A|Residues: 'SHLSNEIPA', 1784-1795 <THO>
 A|Cross-references: EMBL:U07271; NID:g9459665; PIDN:AAA16788.1; PID:g9459666
 C|Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
 C|Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repe
 C|Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F.1-18/Domain: signal sequence #status predicted <SIG>
 F.3-1955/Product: agrin #status predicted <MAT>
 F.39-1783; 1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
 F.39-1647; 1652-1783; 1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
 F.77-126/Domain: Kazal proteinase inhibitor homology <KP11>
 F.152-201/Domain: Kazal proteinase inhibitor homology <KP12>
 F.125-273/Domain: Kazal proteinase inhibitor homology <KP13>
 F.295-344/Domain: Kazal proteinase inhibitor homology <KP14>
 F.370-418/Domain: Kazal proteinase inhibitor homology <KP15>
 F.435-483/Domain: Kazal proteinase inhibitor homology <KP16>
 F.500-548/Domain: Kazal proteinase inhibitor homology <KP17>
 F.588-633/Domain: Kazal proteinase inhibitor homology <KP18>
 F.675-726/Domain: laminin-type EGF-like homology <LE1>
 F.729-773/Domain: laminin-type EGF-like homology <LE2>
 F.801-851/Domain: Kazal proteinase inhibitor homology <KP19>
 F.856-995/Region: serine/threonine-rich
 F.1150-1219/Region: serine/threonine-rich
 F.1233-1264/Domain: EGF homology <EG1>
 F.1294-1448/Domain: laminin G repeat homology <LG1>
 F.1429-1431/Region: motor neuron attachment (L-R-E) motif
 F.1450-1482/Domain: EGF homology <EG2>

F.1489-1521/Domain: EGF homology <EG3>
 F.1560-1711/Domain: laminin G repeat homology <LG2>
 F.1718-1751/Domain: EGF homology <EG4>
 F.1803-1955/Domain: laminin G repeat homology <LG3>
 F.86-105; 94-126; 160-180; 169-201; 233-252; 241-273; 304-323; 312-344; 378-397; 386-418; 443-462;
 1482; 1489-1500; 1494-1510; 1512-1521/Duplicate bonds: #status predicted
 F.390; 659; 764; 814/Binding site: carbohydrate (Aam) (covalent) #status predicted

Query Match 44.8% Score 47; DB 1; Length 1955;
 Best Local Similarity 69.2%; Pred. No. 1.2e+02;
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 8 RNIEDALNQLFRN 20
 |||:|||||:
 Db 1063 RSISALDELFRN 1075

RESULT 4
 574703
 hypothetical protein slr1301 - Synecchocystis sp. (strain PCC 6803)
 C|Species: Synecchocystis sp.
 A|Variety: PCC 6803
 C|Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C|Accession: S74703
 R|Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
 DNA Res. 3, 109-136, 1996
 A|Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
 S.
 A|Reference number: S74322; MUID:97061201; PMID:8905231
 A|Accession: S74703
 A|Status: nucleic acid sequence not shown; translation not shown
 A|Molecule type: DNA
 A|Residues: 1-535 <KAN>
 A|Cross-references: UNIPROT:P72839; EMBL:D90901; GB:AB001339; NID:g1651897; PIDN:BAAL6854
 A|Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C|Genetics:
 A|Start codon: GTG

Query Match 43.8% Score 46; DB 2; Length 535;
 Best Local Similarity 43.8%; Pred. No. 45;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 TNYQRNKNEDALNQL 16
 |||:|||||:
 Db 481 TNYQRNKNEDALNQL 496

RESULT 5
 743060
 agrin - electric ray (Discopyge ommata) (fragment)
 C|Species: Discopyge ommata
 C|Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
 C|Accession: T43060
 R|Smith, M.A.; Magill, S.C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMahar
 submitted to the EMBL Data Library, September 1992
 A|Reference number: Z22308
 A|Accession: T43060
 A|Status: preliminary; translated from GB/EMBL/DBJ
 A|Molecule type: mRNA
 A|Residues: 1-1328 <SMI>
 A|Cross-references: UNIPROT:Q90404; EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.
 C|Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
 C|Keywords: glycoprotein; neuromuscular junction

Query Match 43.8% Score 46; DB 2; Length 1328;
 Best Local Similarity 69.2%; Pred. No. 1.2e+02;
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 8 RNIEDALNQLFRN 20
 |||:|||||:
 Db 448 RSIEALDELFRN 460

RESULT 6
D84647
hypothetical protein At2g25350 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C/Accession: D84647
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentol, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: D84647
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-728 <STO>
A/Cross-references: GB:AE002093; NID:96598555; PIDN:AAD20696.2; GSPDB:GN00139
A/Genetic: A:2g25350
A/Map position: 2

Query Match 42.9%; Score 45; DB 2; Length 728;
Best Local Similarity 64.3%; Pred. No. 88;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TNYQRKNEDALNQLF 14
| | | | | | | | | | | | | | | | | |
Db 386 TTKQRKNEDALNQLF 399

RESULT 7
G81826
hypothetical protein NMA1978 [imported] - Neisseria meningitidis (strain Z249) serogroup
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C/Accession: G81826
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
J.; Holtz, J.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A81775; MUID:2022556; PMID:10761919
A/Accession: G81826
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-56 <PAR>
A/Cross-references: UNIPROT:Q9J743; GB:AL162757; GB:AL157959; NID:97380371; PIDN:CAB8519
A/Experimental source: serogroup A, strain Z2491
C/Genetic: A:Gene: NMA1978

Query Match 41.9%; Score 44; DB 2; Length 56;
Best Local Similarity 46.7%; Pred. No. 8.6;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 ORKRNIEDALNQLF 18
: | | | | | | | | | | | | | | | | | |
Db 28 RRRKRYDNNANSIF 42

RESULT 8
B72386
hypothetical protein - Thermotoga maritima (strain MSB8)
C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C/Accession: B72386
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A/Reference number: A72200; MUID:99287316; PMID:10360571
A/Accession: B72386

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-420 <ARN>
A/Cross-references: UNIPROT:Q9WY19; GB:AE001716; GB:AE000512; NID:94980853; PIDN:AAD35441
A/Experimental source: strain MSB8
C/Genetic: A:Gene: TM0354

Query Match 41.9%; Score 44; DB 2; Length 420;
Best Local Similarity 57.1%; Pred. No. 70;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YQRKNIEDALNQLF 16
| | | | | | | | | | | | | | | | | |
Db 318 YQEVYNNIENSLNQ 331

RESULT 9
G82526
ATP-dependent DNA helicase XP2680 [imported] - Xylella fastidiosa (strain 9a5c)
C/Species: Xylella fastidiosa
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C/Accession: G82526
R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A/Reference number: A82515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: G82526
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-680 <SIM>

A/Cross-references: UNIPROT:Q9PA40; GB:AE004074; GB:AE003849; NID:9107918; PIDN:AAF8547;
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briano, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H.
A/Neco, E.; Docena, C.; El-Porty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, U.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E.
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanelli, R.V.; Sawasaki
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Ze
A/Reference number: A59328
A/Contents: annotation
C/Genetic: A:Gene: XP2680
A/Status: preliminary

Query Match 41.9%; Score 44; DB 2; Length 680;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 NYQRKNIED 12
: | | | | | | | | | | | | | | | | | |
Db 518 SYQRKNIED 528

RESULT 10
F70352
conserved hypothetical protein aq_585 - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C/Accession: F70352
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:98196666; PMID:9537320
A/Accession: F70352

```

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-705 <AQP>
A:Cross-references: UNIPROT:O66850; GB:AEO00697; NID:g2983212; PIDN:AAC06820.1; PID:g298
A:Experimental source: strain VFS
A:Genetics:
A:Gene: aq_585

Query March      41.9% Score 44; DB 2; Length 705;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

3 YQNRKNRNIEDALNQLFNN 20
|:::|||||
Db          303 YFKRKRDLEIALSRUKEN 320

RESULT 11
E90016
hyaluronate lyase precursor [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E90016
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Hoi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: AB9758; MUID:21311952; PMID:11418146
A:Accession: E90016
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-809 <KUR>
A:Cross-references: UNIPROT:Q9S657; GB:BA000018; PID:g13702002; PIDN:BAB43294.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: hyaA

Query Match      41.9% Score 44; DB 2; Length 809;
Best Local Similarity 42.1%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

2 NYQRKNRIEDALNQLFNN 20
|:::|||||
Db          136 NTDENKKVKDLEMWLRN 154

RESULT 12
D70128
conserved hypothetical protein BB0228 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: D70128
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Claydon, R.; Lathigra, R.; White, B.; Peterson, J.; Karlavag, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugr, B.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 589-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: D70128
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-971 <KLE>
A:Cross-references: UNIPROT:O51246; GB:AB001133; GB:AB000783; NID:g2688120; PIDN:AAC6662
A:Experimental source: strain BJ1

Query Match      41.9% Score 44; DB 2; Length 971;
Best Local Similarity 47.4%; Pred. No. 1.7e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

2 NYQRKNRIEDALNQLFNN 20
|:::|||||

```

```

DB          896  SYRRKMLNISDLSRODIRN  914

RESULT 13
T19481
hypochemical protein C26C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T19481; T25434
R:McMurray, A.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z19128
A:Accession: T19481
A:Status: preliminary; translated from GB/EMBL/DBDJI
A:Molecule type: DNA
A:Residues: 1-1879 <WIL>
A:Cross-references: UNIPROT:Q18210; EMBL:Z72503; PIDN:CAA96600.1; GSPDB:GN00019; CESP:C26C6
A:Experimental source: clone C26C6
R:McMurray, A.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20033
A:Accession: T25434
A:Status: preliminary; translated from GB/EMBL/DBDJI
A:Molecule type: DNA
A:Residues: 1-1879 <W12>
A:Cross-references: EMBL:Z72517; PIDN:CAA96697.1; GSPDB:GN00019; CESP:C26C6.1
A:Experimental source: clone T28P4
C:Genetics:
A:Gene: CESP:C26C6.1
A:Map position: 1
A:Introns: 42/3; 295/2; 375/2; 489/3; 665/2; 980/3; 1233/2; 1292/1; 1325/3; 1349/3; 1486/3
F:368-422/Domain: bromodomain homology <BRO1>
F:560-615/Domain: bromodomain homology <BRO2>
F:737-792/Domain: bromodomain homology <BRO3>

Query Match          41.9%; Score 44; DB 2; Length 1879;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB          692  NIQOHQSRMEDALIMOL  707

RESULT 14
H64218
preprotein translocase secY - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 09-Jul-2004
C:Accession: H64218
R:Risser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhman, J.; Nguyen, D.; Uitterlacker, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.R.;
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: H64218
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <TIGR>
A:Cross-references: UNIPROT:P47416; GB:U39695; GB:L43967; NID:g1045833; PID:g1045854; TIGR:
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: preprotein translocase secY

Query Match          41.4%; Score 43.5; DB 2; Length 475;
Best Local Similarity 38.5%; Pred. No. 95;
Matches 10; Conservative 5; Mismatches 2; Indels 9; Gaps 1;

DB          439  TNYOOLRRNLAIIEVOTAKODSLLEOL  464

1 TNYORKNRNI-----EDALNOL 17
||||: |||:

```

RESULT 15

G71611

hypothetical protein PFB0560w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: G71611

R: Gardner, M.J.; Tetteijn, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.: Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; WUID:99021743; PMID:9804551

A:Accession: G71611

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2206 <GAR>

A:Cross-references: GB:AE001403; GB:AE001362; NID:G3845216; PIDN:AACT1901.1; PID:G384521

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0560w

Query Match 41.4%; Score 43.5; DB 2; Length 2206;

Best Local Similarity 36.8%; Pred. No. 4.7e+02;

Matches 7; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

QY 2 NYQRNKRNT-DALNQLFR 19

Db 1972 NYNNNNNNVDVYLNKIKYK 1990

Search completed: October 27, 2005, 18:57:25
Job time : 14.3721 secs

This Page Blank (intentional)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 27, 2005, 17:34:49 ; Search time 58.1395 Seconds
(without alignments)
176.155 Million cell updates/sec

Title: US-10-612-090-15
Perfect score: 105
Sequence: 1 TNYQRNKNIEDALNQLFRN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	1148	2 Q9H7S7	Q9H7S7 homo sapien
2	105	100.0	6995	2 Q96RK2	Q96RK2 homo sapien
3	105	100.0	22152	2 Q8WY17	Q8WY17 homo sapien
4	79	75.2	258	2 Q9DHI1	Q9DHI1 mus musculu
5	54	51.4	482	2 Q7QDT5	Q7QDT5 anopheles g
6	52	49.5	390	2 Q7RCC0	Q7RCC0 plasmodium
7	51	48.6	841	2 Q9S6T8	Q9S6T8 streptomyce
8	48	45.7	109	2 Q9XZ81	Q9XZ81 plasmodium
9	48	45.7	132	2 Q6FCV7	Q6FCV7 acinetobact
10	48	45.7	319	2 Q8ZC88	Q8ZC88 salmonella
11	48	45.7	360	2 Q7SB96	Q7SB96 neurospora
12	48	45.7	666	2 Q32495	Q32495 bacteroides
13	48	45.7	708	2 Q9GQ16	Q9GQ16 manduca sex
14	48	45.7	797	2 Q8KH33	Q8KH33 micromonos
15	48	45.7	841	2 Q8ZEB8	Q8ZEB8 streptomyce
16	48	45.7	956	2 Q6LCU0	Q6LCU0 photobacter
17	48	45.7	2879	2 Q6LFO9	Q6LFO9 plasmodium
18	47	44.8	349	2 Q7NFX9	Q7NFX9 gloeobacter
19	47	44.8	447	2 Q6W3Q4	Q6W3Q4 alvineella p
20	47	44.8	575	2 Q812A1	Q812A1 plasmodium
21	47	44.8	1048	2 Q8KJ11	Q8KJ11 rhicobium l
22	47	44.8	1955	1 AGRN_CHICK	P31696 gallus gall
23	46.5	44.3	2060	2 Q7Q3S9	Q7Q3S9 anopheles g
24	46	43.8	109	2 Q9NHJ1	Q9NHJ1 plasmodium
25	46	43.8	109	2 Q9NHL5	Q9NHL5 plasmodium
26	46	43.8	171	1 PACA_RANRI	Q09169 r glucagon-
27	46	43.8	376	2 Q8PMN7	Q8PMN7 methanosarc
28	46	43.8	535	2 P72839	P72839 synecocyst
29	46	43.8	921	2 Q7RP27	Q7RP27 plasmodium
30	46	43.8	970	2 Q7RG16	Q7RG16 plasmodium
31	46	43.8	1049	2 Q83611	Q83611 ectromelia

32	46	43.8	1113	2 Q8JL89	Q8JL89 ectromelia
33	46	43.8	1328	1 AGRN_DISOM	Q90404 discopyge o
34	46	43.8	2760	2 Q815Y2	Q815Y2 plasmodium
35	45.5	43.3	792	2 Q6R230	Q6R230 debaryomyce
36	45.5	43.3	992	2 Q99877	Q99877 trypansomera
37	45	42.9	109	2 Q8T4V4	Q8T4V4 plasmodium
38	45	42.9	110	2 Q931B6	Q931B6 staphylococ
39	45	42.9	247	2 Q7P8Y8	Q7P8Y8 anopheles g
40	45	42.9	248	2 Q81KY4	Q81KY4 plasmodium
41	45	42.9	314	2 Q8MTT3	Q8MTT3 tetrahymena
42	45	42.9	315	2 Q9ZTCS	Q9ZTCS listeria mo
43	45	42.9	316	2 Q71W10	Q71W10 listeria mo
44	45	42.9	386	2 Q81DW4	Q81DW4 plasmodium
45	45	42.9	447	2 Q6W3M3	Q6W3M3 alvineella p

ALIGNMENTS

RESULT 1
Q9H7S7 PRELIMINARY; PRT; 1148 AA.
ID Q9H7S7
AC Q9H7S7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ14303.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta; DOI=10.1038/ng1285;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosobata T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hoshino T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiyasu S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihara K., Matsumura H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Motiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Itakura S., Fukumitsu Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujimura T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunaga K., Nakajima Y., Mizuno T., Moritaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK024365; BAB14899.1; -.
DR HSSP; Q9DHI1; IIVZ.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 7.
DR PROSITE; PS0024; SEA; 3.
SQ SEQUENCE 1148 AA; 127957 MW; 3861B0D5EFDFA8C CRC64;

```

Query Match          100.0%; Score 105; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TNYQNRKNRINEDALNQLFRN 20
    |||||
    985 TNYQNRKNRINEDALNQLFRN 1004

RESULT 2
ID Q96RK2 PRELIMINARY; PRT; 6995 AA.
AC Q96RK2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Mucin 16 (Fragment).
GN Name=MUC16;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347965; PubMed=11369781; DOI=10.1074/jbc.M103554200;
RA Yin B.W., Lloyd K.O.;
RT "Molecular cloning of the cal25 ovarian cancer antigen. Identification
RT as a new mucin, muc16."
RL J. Biol. Chem. 276:27371-27375 (2001).
[2]
RP SEQUENCE FROM N.A.
RX LLOYD K.O., Yin B.W.T.;
RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF361486; AAK74120.3; -.
DR HSSP; Q9DIH1; 11VZ.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 20.
DR SMART; SMO0200; SEA; 10.
DR PROSITE; PS00152; ATPase_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00024; SEA; 6.
RT NON TER 1 1
SO SEQUENCE 6995 AA; 744958 MW; 80C797DBDF33A2B CRC64;

Query Match          100.0%; Score 105; DB 2; Length 6995;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TNYQNRKNRINEDALNQLFRN 20
    |||||
    6832 TNYQNRKNRINEDALNQLFRN 6851

Db 6832 TNYQNRKNRINEDALNQLFRN 6851

RESULT 3
ID Q8WX17 PRELIMINARY; PRT; 22152 AA.
AC Q8WX17;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Ovarian cancer related tumor marker CA125.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21646939; PubMed=11786729; DOI=10.1159/000050638;
RA O'Brien T.J., Beard J.B., Underwood L.J., Dennis R.A., Santin A.D.,
RA York L.;
RT "The CA 125 gene: an extracellular superstructure dominated by repeat
RT sequences."
RL Tumour Biol. 22:348-366 (2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA O'Brien T.J., Underwood L.J., Beard J.B.;
RL Submitted (OCT-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF414442; AAL65133.2; -.
DR Genew; HGNC:15582; MUC16.
DR Pfam; PF01390; SEA; 51.
DR SMART; SMO0200; SEA; 23.
DR PROSITE; PS00152; ATPase_ALPHA_BETA; UNKNOWN_2.
DR PROSITE; PS00024; SEA; 11.
SQ SEQUENCE 22152 AA; 235268 MW; B3E7BDF1997A440 CRC64;

Query Match          100.0%; Score 105; DB 2; Length 22152;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TNYQNRKNRINEDALNQLFRN 20
    |||||
    21989 TNYQNRKNRINEDALNQLFRN 22008

Db 21989 TNYQNRKNRINEDALNQLFRN 22008

RESULT 4
ID Q9DIH1 PRELIMINARY; PRT; 258 AA.
AC Q9DIH1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched
DE library, clone:111000814 product:hypothetical SEA domain containing
DE protein, full insert sequence.
GN Name=111000814R1k;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carinici P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44 (1999).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
[4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carinici P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630 (2000).
[5]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20530913; PubMed=11076661; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carinici P.,
RA Kono H., Akiyama J., Nishi K., Kitsumi T., Tashiro H., Itoh M.,

```

```

DR InterPro; IPR001320; Ion_glu_receptor.
DR InterPro; IPR001622; K-channel_pore.
DR InterPro; IPR001508; NMDA_receptor.
DR InterPro; IPR001311; SBP_glu_receptor.
DR Pfam; PF00060; Lig_chan; 1.
DR PRINTS; PR00177; NMDARECEPTOR.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 482 AA; 54936 MW; 2BB7187196C8DF7 CRC64;

Query Match 51.4%; Score 54; DB 2; Length 482;
Best Local Similarity 52.6%; Pred. No. 11;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 NYQNRKNEDALNQLFRN 20
   :| | | | | | | | | |
Db 272 DYFRNMKAEDVLYQWNRN 290

RESULT 6
ID Q7RCO PRELIMINARY; PRT; 390 AA.
AC Q7RCO;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY05864;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
NCBI Taxid=73239;

```

RP	SEQUENCE FROM N.A.
RC	STRAIN=17XN1;
RK	PUBMED=12368865; DOI=10.1038/nature01099;
RA	Catillon J.M., Angiuoli S.V., Sun B.B., Kooij T.W., Perteau M.,
RA	Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA	Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA	Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA	Cho J.K., Quackenbush J., Sedegah M., Shoathi A., Cummings L.M.,
RA	Florens L., Yates F.R. III, Raine J.D., Sindren R.E., Harris M.A.,
RA	Cunningham D.A., Preiser P.R., Bergman L.W., Vaitya A.B.,
RA	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA	Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA	Carucci D.J.;
RT	"Genome sequence and comparative analysis of the model rodent malaria
RT	parasite Plasmodium yoelii yoelii."
RL	Nature 419:512-519(2002).
CC	-I- CAUTION: The sequence shown here is derived from an
CC	EMBL/Genebank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
DR	EMBL; AABL01001920; EAA17959.1; -.
DR	InterPro; IPR0011591; Botulinum.
DR	ProDom; PD001963; Botulinum; 1.
KM	Hypothetical protein.
SO	SEQUENCE 390 AA; 45937 MW; 0FA3B21E0A68BA CRC64;
OY	2 NYORNRKINEDALNQ 16
Db	233 NYEKNNENIMENIQ 247
RESULT 7	
ID	Q9S6T8
AC	Q9S6T8 PRELIMINARY; PRG; 841 AA.
DT	01-MAY-2000 (TREMBLErel. 13, Created)
DT	01-MAY-2000 (TREMBLErel. 13, Last sequence update)
DT	01-OCT-2003 (TREMBLErel. 25, Last annotation update)

```

DE Putative C1P-family ATP-binding protease.
GN ORFNames=SCE94.24c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitsch B., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RN Nature 417:141-147(2002).
RX EMBL: AL939116; CAB40873.1; -.
DR PIR: T36384; T36384.
DR HSSP: P03815; IJBK.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0004518; F:nuclease activity; IEA.
DR GO: GO:0000166; F:nucleotide binding; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0005515; F:protein binding; IEA.
DR GO: GO:0006289; P:nucleotide-excision repair; IEA.
DR GO: GO:0019538; P:protein metabolism; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001270; Chaprtn_C1Pa/B.
DR InterPro: IPR004176; C1P_N.
DR InterPro: IPR001943; UVRB/C.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF02861; C1P_N; 2.
DR Pfam: PF02151; UVR_1.
DR PRINTS: PR00300; CLPROTEASEA.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS50151; UVR; 1.
DR ATP-binding; Complete proteome; Protease.
SQ SEQUENCE 841 AA; 93013 MW; 101BE09AADA7FFD CRC64;

Query Match 48.6%; Score 51; DB 2; Length 841;
Best Local Similarity 47.4%; Pred. No. 57;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 TNYQNRKNIEDALNQLFR 19
DB 685 SNYERMKVKVQDELKQHR 703

RESULT 8
OYXZ81 PRELIMINARY; PRT; 109 AA.
AC O9XZ81;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE DBL alpha protein (Fragment).
GN Name=var;
OS Plasmodium falciptarum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=20078856; PubMed=10613695; DOI=10.1016/S0166-6851(99)00159-0;
RA Taylor H.M., Kyes S.A., Harris D., Kriek N., Newbold C.I.;

```

```

RT "A study of var gene transcription in vitro using universal var gene
RT primers.";
RN Mol. Biochem. Parasitol. 105:13-23(2000).
RX EMBL: AF133853; AAD33614.1; -.
DR NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 109 AA; 13027 MW; 7EFD9E0BD14AE174 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 109;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 3 YQNRKNIEDALNQLFRN 20
DB 15 HQQRKRLKENLKQMFEN 32

RESULT 9
OYFCV7 PRELIMINARY; PRT; 132 AA.
AC O6FCV7;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=ACIAD1226;
OS Acinetobacter sp. (strain ADP1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=62977;
RN [1]
RP SEQUENCE FROM N.A.
RA Barbe V., Vallenet D., Fonknechten N., Kreilmeyer A., Ozas S.,
RA Labarre L., Crueviller S., Robert C., Duprat S., Wincker P.,
RA Orston L.N., Weissenbach J., Marliere P., Cohen G.N., Medigue C.;
RT "Unique features revealed by the genome sequence of Acinetobacter sp.
RT ADP1, a versatile and naturally transformation competent bacterium.";
RL Nucleic Acids Res. 0:0-0(2004).
DR EMBL: CR543861; CAG68102.1; -.
KW Complete proteome.
SQ SEQUENCE 132 AA; 14775 MW; E8B481E44AB30F2A CRC64;

Query Match 45.7%; Score 48; DB 2; Length 132;
Best Local Similarity 56.2%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 NYQNRKNIEDALNQL 17
DB 80 DYMPNKKPBDALNKL 95

RESULT 10
OYQZC8 PRELIMINARY; PRT; 319 AA.
AC O8ZC8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Homologous to secreted protein sopp.
GN OrderedLocNames=STM0972;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L72;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;

```



```

DT 01-OCT-2002 (TRENBLrel. 22, last sequence update)
DE 01-MAR-2004 (TRENBLrel. 26, last annotation update)
DR CaLR4.
GN Name=calR4;
OS Micromonospora echinospora (Micromonospora purpurea).
OC Bacteria; Actinobacteri; Actinobacteridae; Actinomycetales;
CX Micromonosporaceae; Micromonosporaceae; Micromonospora.
OK NCBI_TaxID=1877;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 15839;
RX MEDLINE=22171414; PubMed=12183629; DOI=10.1126/science.1072105;
RA Ahlert J., Shepard E., Lomovskaya N., Zaopoulos E., Stafila A.,
  Bachmann B.O., Huang K., Fongstein L., Czisny A., Whitlam R.E.,
  Farnet C.M., Thorson J.S.;
  "The calicheamicin gene cluster and its iterative type I enediyne
  PKs.";
  Science 297:1173-1176(2002).
RL EMBL; AF497482; AAA94782.1; -.
DR HSSP; P03815; IUBK.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:000166; F:nucleotide binding; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR GO; GO:0019538; F:protein metabolism; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001270; Chaprinin_c1pA/B.
DR InterPro; IPR004176; Clp_N.
DR InterPro; IPR001943; UvrB/C.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS50151; UVR; 1.
KW ATP-binding.
SQ SEQUENCE 797 AA; 88267 MW; 9E107A6CA894C190 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 797;
Best Local Similarity 47.4%; Pred. No. 1.5e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 TNYQRNKRNIEDALNQLFR 19
DB 648 SNYDRMKQKVNDELKQHR 666

RESULT 15
O82EB8 PRELIMINARY; PRT; 841 AA.
AC O82EB8;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, last annotation update)
DE Putative ATP-dependent Clp protease.
GN Name=clpC2; OrderedlocusNames=SAV4697;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteri; Actinobacteridae; Actinomycetales;
CX Streptomycinase; Streptomycetaceae; Streptomycetes.
OK NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shida T.,
  Sakaki Y., Hattori M., Omura S.;
  "Complete genome sequence and comparative analysis of the industrial
  microorganism Streptomyces avermitilis.";
  Nat. Biotechnol. 21:526-531(2003).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.21143198;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
  Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
  Kikuchi H., Shida T., Sakaki Y., Hattori M.,
  "Genomic sequence of an industrial microorganism Streptomyces
  avermitilis: deducing the ability of producing secondary
  metabolites.";
  Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RL EMBL; AP005040; BAC72409.1; -.
DR HSSP; P03815; IUBK.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004518; F:nuclease activity; IEA.
DR GO; GO:000166; F:nucleotide binding; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0006289; F:nucleotide-excision repair; IEA.
DR GO; GO:0019538; F:protein metabolism; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001270; Chaprinin_c1pA/B.
DR InterPro; IPR004176; Clp_N.
DR InterPro; IPR001943; UvrB/C.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS50151; UVR; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 841 AA; 93023 MW; 265866F0CA9BCCF7 CRC64;

Query Match 45.7%; Score 48; DB 2; Length 841;
Best Local Similarity 47.4%; Pred. No. 1.6e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 TNYQRNKRNIEDALNQLFR 19
DB 685 SNYERMKKVNDELKQHR 703

```

Search completed: October 27, 2005, 18:55:10
 Job time : 61.1395 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:31:33 ; Search time 72.6698 Seconds
(without alignments)
111.765 Million cell updates/sec

Title: US-10-612-090-16

Perfect score: 107

Sequence: 1 FRNSSIKSYFSDQVSTFRSV 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseq1980s:*
2: geneseq1990s:*
3: geneseq2000s:*
4: geneseq2001s:*
5: geneseq2002s:*
6: geneseq2003as:*
7: geneseq2003bs:*
8: geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	21	ADJ66743	Adj66743 Human Muc
2	107	100.0	108	ADJ66729	Adj66729 Human Muc
3	107	100.0	108	ADJ66740	Adj66740 Human Muc
4	107	100.0	174	ADM30756	Adm30756 Human can
5	107	100.0	178	ABP30979	Abp30979 Predicted
6	107	100.0	178	ADA08642	Ada08642 Human O77
7	107	100.0	178	ADF08985	Adf08985 Secreted
8	107	100.0	233	ABP30978	Abp30978 Extracell
9	107	100.0	233	ADA08641	Ada08641 Human O77
10	107	100.0	223	ADF08984	Adf08984 Secreted
11	107	100.0	284	ABU54859	Abu54859 Human CA1
12	107	100.0	318	ABP31025	Abp31025 Antio aci
13	107	100.0	318	ADA08747	Ada08747 Human O77
14	107	100.0	318	ADF09090	Adf09090 Secreted
15	107	100.0	367	AAE12631	Aae12631 Human gen
16	107	100.0	438	AAE12555	Aae12555 Human ova
17	107	100.0	438	AAE12623	Aae12623 Human gen
18	107	100.0	438	ABP30973	Abp30973 Partial p
19	107	100.0	438	ABP30899	Abp30899 0772P c10
20	107	100.0	438	ADA08636	Ada08636 Human O77
21	107	100.0	438	ADA08543	Ada08543 Human ova
22	107	100.0	438	ADP08886	Adp08886 Secreted
23	107	100.0	438	ADP08979	Adp08979 Secreted
24	107	100.0	438	ADG46173	Adg46173 Human ova
25	107	100.0	439	ABU54709	Abu54709 Human CA1

26	107	100.0	526	AAE12634	Aae12634 Human gen
27	107	100.0	748	ADS94302	Ads94302 CA 125/O7
28	107	100.0	809	ADS94303	Ads94303 CA 125/O7
29	107	100.0	833	AAE12554	Aae12554 Human ova
30	107	100.0	833	ABP30898	Abp30898 0772P c10
31	107	100.0	833	ADA08542	Ada08542 Human ova
32	107	100.0	833	ADP08885	Adp08885 Secreted
33	107	100.0	833	ADG46172	Adg46172 Human ova
34	107	100.0	914	AAE12552	Aae12552 Human ova
35	107	100.0	914	AAE12553	Aae12553 Human ova
36	107	100.0	914	ABP30968	Abp30968 Hypothetcl
37	107	100.0	914	ABP30896	Abp30896 0772P pro
38	107	100.0	914	ADA08631	Ada08631 Human O77
39	107	100.0	914	ADA08465	Ada08465 Human ova
40	107	100.0	914	ADF08974	Adf08974 Secreted
41	107	100.0	914	ADP08808	Adp08808 Secreted
42	107	100.0	914	ADG46095	Adg46095 Human ova
43	107	100.0	914	ADN40451	Adn40451 Human bre
44	107	100.0	1148	AAB95836	Aab95836 Human pro
45	107	100.0	1148	ABB50283	Abb50283 HOST-1 cv

ALIGNMENTS

RESULT 1

ADJ66743
ID ADJ66743 standard; peptide; 21 AA.

XX ADJ66743;

XX 06-MAY-2004 (first entry)

XX	Human Muc1/Muc16 protein-related peptide SegID16.
XX	monoclonal antibody; epitope; non-shed extracellular portion;
XX	shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
XX	breast cancer; ovarian cancer.
XX	
OS	Homo sapiens.
XX	
XX	WO2004005470-A2.
PN	
XX	
PD	15-JAN-2004.
XX	
PF	03-JUL-2003; 2003WO-US020907.
XX	
XX	03-JUL-2002; 2002US-0393094P.
PR	
XX	(IMMU-) IMMUNOGEN INC.
PA	
XX	Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
PI	Water CA;
XX	
PI	WPI; 2004-091350/09.
DR	
XX	
PT	New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
PT	treating or monitoring malignancies, such as breast or ovarian cancer.
XX	
PS	Claim 14; SEQ ID NO 16; 113pp; English.
XX	
CC	This invention relates to a novel isolated monoclonal antibody that
CC	specifically binds to an epitope of a non-shed extracellular portion of a
CC	shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC	useful for the development of compounds with a cytostatic or for gene
CC	therapy. The composition and methods are useful in diagnosing, treating
CC	or monitoring malignancies, such as breast or ovarian cancer. The present
CC	sequence is that of a peptide which was used for raising antibodies to
CC	the extracellular, non-shed region of Muc16 and which is claimed in the
CC	specification.
XX	
XX	
SQ	Sequence 21 AA;

```

Query Match          100.0%; Score 107; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 2,6e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FRNSIKSYFSDCOVSTFRSV 21
        |||||
        1 FRNSIKSYFSDCOVSTFRSV 21
DB      18 FRNSIKSYFSDCOVSTFRSV 38

RESULT 2
ADJ66729 ADJ66729 standard; protein; 108 AA.
XX
AC      ADJ66729;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Human Muc16 epitope amino acid sequence.
XX
KW      monoclonal antibody; epitope; non-shed extracellular portion;
KW      shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
KW      breast cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2004005470-A2.
XX
PD      15-JAN-2004.
XX
PF      03-JUL-2003; 2003WO-US020907.
XX
PR      03-JUL-2002; 2002US-0393094P.
XX
PA      (IMMU-) IMMUNOGEN INC.
XX
PI      Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,
PI      Vater CA;
XX
DR      WPI; 2004-091350/09.
XX
PT      New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
PT      treating or monitoring malignancies, such as breast or ovarian cancer.
XX
PS      Claim 13; SEQ ID NO 2; 113pp; English.
XX
CC      This invention relates to a novel isolated monoclonal antibody that
CC      specifically binds to an epitope of a non-shed extracellular portion of a
CC      shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC      useful for the development of compounds with a cytostatic or for gene
CC      therapy. The composition and methods are useful in diagnosing, treating
CC      or monitoring malignancies, such as breast or ovarian cancer. The present
CC      sequence is that of a Muc epitope of the invention.
XX
SQ      Sequence 108 AA;

Query Match          100.0%; Score 107; DB 8; Length 108;
Best Local Similarity 100.0%; Pred. No. 1,7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FRNSIKSYFSDCOVSTFRSV 21
        |||||
        1 FRNSIKSYFSDCOVSTFRSV 21
DB      18 FRNSIKSYFSDCOVSTFRSV 38

RESULT 3
ADJ66740 ADJ66740 standard; protein; 108 AA.
XX
AC      ADJ66740;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Human Muc16 GST fusion protein amino acid sequence SeqID13.
XX

```

```

XX      monoclona antibody; epitope; non-shed extracellular portion;
KW      shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
KW      breast cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2004005470-A2.
XX
PD      15-JAN-2004.
XX
PF      03-JUL-2003; 2003WO-US020907.
XX
PR      03-JUL-2002; 2002US-0393094P.
XX
PA      (IMMU-) IMMUNOGEN INC.
XX
PI      Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,
PI      Vater CA;
XX
DR      WPI; 2004-091350/09.
XX
PT      New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
PT      treating or monitoring malignancies, such as breast or ovarian cancer.
XX
PS      Disclosure; SEQ ID NO 13; 113pp; English.
XX
CC      This invention relates to a novel isolated monoclonal antibody that
CC      specifically binds to an epitope of a non-shed extracellular portion of a
CC      shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC      useful for the development of compounds with a cytostatic or for gene
CC      therapy. The composition and methods are useful in diagnosing, treating
CC      or monitoring malignancies, such as breast or ovarian cancer. The present
CC      sequence is that of the human Muc16 GST fusion protein which was used in
CC      the exemplification of the invention.
XX
SQ      Sequence 108 AA;

Query Match          100.0%; Score 107; DB 8; Length 108;
Best Local Similarity 100.0%; Pred. No. 1,7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FRNSIKSYFSDCOVSTFRSV 21
        |||||
        1 FRNSIKSYFSDCOVSTFRSV 38
DB      18 FRNSIKSYFSDCOVSTFRSV 38

RESULT 4
ADM30756 ADM30756 standard; protein; 174 AA.
XX
AC      ADM30756;
XX
DT      20-MAY-2004 (first entry)
XX
DE      Human cancer linked protein referred to as 192292 SeqID 4.
XX
KW      human; cancer-related gene; screening assay; immuno-conjugate;
KW      cytostatic; immunostimulant; cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2003075854-A2.
XX
PD      18-SEP-2003.
XX
PF      07-MAR-2003; 2003WO-US007147.
XX
PR      07-MAR-2002; 2002US-0362527P.
XX
PA      (AVAL-) AVALON PHARM.
XX
PI      Young PE, Ebner R, Strovvel JW;
XX

```

XX WPI; 2003-748320/70.
DR N-PSDB; ADM30755.
XX
PT Identifying an agent that modulates the activity of any of three specific
PT cancer-related genes, potentially useful in treating (ovarian) cancer,
PT comprises detecting a difference in expression of the gene in the
PT presence of the agent.
XX
PS Claim 11; SEQ ID NO 4; 57bp; English.
XX
CC This invention relates to a novel method of identifying an agent that
CC modulates the activity of a cancer-related gene. Specifically, it refers
CC to a screening assay to identify potential antitumor agents, as well as
CC methods to assess the cancerous state of a cell. The present invention
CC describes antibodies against the expression products of these cancer-
CC related genes that are capable of targeting a cancerous cell in vivo.
CC Furthermore it provides immuno-conjugates containing such antibodies,
CC which can be used to deliver target therapeutics thereto. Accordingly,
CC these compositions that exhibit cytostatic and immunostimulant activities
CC are useful for preventing or treating cancer in humans and in particular
CC the treatment of ovarian cancer. This polypeptide sequence is a protein
CC encoded by a human cancer-linked gene of the invention.
XX
SQ Sequence 174 AA;
XX
Query Match 100.0%; Score 107; DB 7; Length 174;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FRNSSIKSYFSDCOVSTFRSV 21
|||
28 FRNSSIKSYFSDCOVSTFRSV 48
DB
RESULT 5
ABP30979
ID ABP30979 standard; protein; 178 AA.
XX
AC ABP30979;
XX
DT 02-JUL-2002 (first entry)
XX
DE Predicted extracellular domain of 0772P.
XX
XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
OS Homo sapiens.
XX
PN WO200206317-A2.
XX
PD 24-JAN-2002.
XX
PF 17-JUL-2001; 2001WO-US022635.
XX
PI 17-JUL-2000; 2000US-00617747.
XX
PR 10-AUG-2000; 2000US-00636801.
XX
PR 20-SEP-2000; 2000US-00667857.
XX
PR 04-APR-2001; 2001US-00827271.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2002-164781/21.
XX
PT Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
PS Example 13; Page 361-362; 408bp; English.

XX
CC This invention relates to polypeptides comprising an immunogenic portion
CC of an ovarian carcinoma protein which acts as an immunostimulant and is
CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations and antigen presenting cells that express
CC the polypeptides are useful for stimulating an immune response in a
CC patient and treating ovarian cancer. This sequence represents protein
CC related to the invention
XX
SQ Sequence 178 AA;
XX
Query Match 100.0%; Score 107; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 FRNSSIKSYFSDCOVSTFRSV 21
|||
87 FRNSSIKSYFSDCOVSTFRSV 107
DB
RESULT 6
ADA08642
ID ADA08642 standard; protein; 178 AA.
XX
AC ADA08642;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human 0772P partial protein #12.
XX
XX Human; gene therapy; ovarian cancer; cancer.
XX
OS Homo sapiens.
XX
PN US2003091580-A1.
XX
PD 15-MAY-2003.
XX
PF 17-JUL-2001; 2001US-00907969.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PI (MITC/) MITCHAM J L.
XX
PI (KING/) KING G E.
XX
PI (ALGA/) ALGATE P A.
XX
PI (FLIN/) FLING S P.
XX
PI (RETT/) RETTER M W.
XX
PI (FANG/) FANGER G R.
XX
PI (REED/) REED S G.
XX
PI (VEDV/) VEDVICK T S.
XX
PI (CART/) CARTER D.
XX
PI (HILL/) HILL P.
XX
PI (ALBO/) ALBONE E.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2003-532352/50.
XX
PT New isolated 0772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
PT cancer.
XX
PS Disclosure; Page 28; 371bp; English.
XX
XX The invention relates to an isolated 0772P polypeptide, which has the
XX structure fully defined in the specification. The composition containing
XX the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
XX or antigen presenting cells are useful for stimulating an immune response
XX and treating ovarian cancer. Detecting the presence of the
XX polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
XX carcinoma cDNAs and protein cDNAs were identified using microarray
XX technology. The present sequence represents a human ovarian carcinoma

CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 107; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 FRNSIKSYFSDCOVSTFRSV 21
 |||||
 DB 87 FRNSIKSYFSDCOVSTFRSV 107
 RESULT 7
 ID ADF08985 standard; protein; 178 AA.
 XX
 AC ADF08985;
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Secreted ovarian carcinoma antigen seqid 489.
 XX
 KW gene therapy; protein therapy; vaccine; antibody inhibition;
 KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
 KW secreted ovarian carcinoma antigen.
 XX
 OS Homo sapiens.
 XX
 PN US2003124140-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 17-JUL-2002; 2002US-00198053.
 XX
 PR 17-DEC-1998; 98US-00215681.
 PR 17-DEC-1998; 98US-00216003.
 PR 23-JUN-1999; 99US-00338933.
 PR 24-SEP-1999; 99US-00404879.
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00657857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 PR 17-JUL-2001; 2001US-00907969.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Bangur CS, Retter MW, Fanger GR, Hill P;
 DR WPI; 2003-897152/82.
 XX
 PT Oncogenic nucleic acids useful for the prevention, diagnosis and
 PT treatment of breast cancer.
 XX
 PS Example 13; SEQ ID NO 489; 399pp; English.
 CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancer. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patient's own production of (II).
 CC Additionally, (I) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigens in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and
 CC antagonists may be used to regulate expression and activity and as

CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunoassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 107; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 FRNSIKSYFSDCOVSTFRSV 21
 |||||
 DB 87 FRNSIKSYFSDCOVSTFRSV 107
 RESULT 8
 ID ABP30978 standard; protein; 233 AA.
 XX
 AC ABP30978;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Extracellular and transmembranal regions of 0772P.
 XX
 KW Human; immunostimulant; cytoslatic; cancer; ovarian carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN WO200206317-A2.
 XX
 PD 24-JAN-2002.
 XX
 PF 17-JUL-2001; 2001WO-US022635.
 XX
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR,
 PI Reed SG, Vedvick TS, Carter D, Hill P, Albone B;
 XX
 DR WPI; 2002-164781/21.
 XX
 PT Polypeptides comprising an immunogenic portion of an ovarian carcinoma
 PT protein or its variants, useful for stimulating an immune response in a
 PT patient and treating ovarian cancer.
 XX
 PS Example 13; Page 361; 408pp; English.
 CC This invention relates to polypeptides comprising an immunogenic portion
 CC of an ovarian carcinoma protein which acts as an immunostimulant and is
 CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations and antigen presenting cells that express
 CC the polypeptides are useful for stimulating an immune response in a
 CC patient and treating ovarian cancer. This sequence represents protein
 CC related to the invention
 CC
 SQ Sequence 233 AA;
 Query Match 100.0%; Score 107; DB 5; Length 233;
 Best Local Similarity 100.0%; Pred. No. 4.1e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 FRNSIKSYFSDCOVSTFRSV 21
 |||||
 DB 87 FRNSIKSYFSDCOVSTFRSV 107

```
RESULT 9
ADA08641
XX ID ADA08641 standard; protein; 233 AA.
XX AC ADA08641;
XX DT 06-NOV-2003 (first entry)
XX DE Human O772P partial protein #11.
XX KW human; gene therapy; ovarian cancer; cancer.
XX OS Homo sapiens.
XX PN US2003091580-A1.
XX PD 15-MAY-2003.
XX PF 17-JUL-2001; 2001US-00907969.
XX PR 18-JUN-2001; 2001US-00884441.
XX PA (MITC/) MITCHAM J L.
XX PA (KING/) KING G E.
XX PA (ALGA/) ALGATE P A.
XX PA (FLIN/) FLING S P.
XX PA (RETT/) RETTER M W.
XX PA (FANG/) FANGER G R.
XX PA (REED/) REED S G.
XX PA (VEDV/) VEDVICK T S.
XX PA (CART/) CARTER D.
XX PA (HILL/) HILL P.
XX PA (ALBO/) ALBONE E.
XX PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
XX PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX DR WPI; 2003-532352/50.
XX PT New isolated O772P polypeptides and polynucleotides, useful in gene
XX PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
XX PT cancer.
XX PS Example 13; Page 28; 371pp; English.
XX CC The invention relates to an isolated O772P polypeptide, which has the
XX CC structure fully defined in the specification. The composition containing
XX CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
XX CC or antigen presenting cells are useful for stimulating an immune response
XX CC and treating ovarian cancer. Detecting the presence of the
XX CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
XX CC carcinoma cDNAs and protein cDNAs were identified using microarray
XX CC technology. The present sequence represents a human ovarian carcinoma
XX CC antigen.
XX SQ Sequence 233 AA;
OY Query Match 100.0%; Score 107; DB 7; Length 233;
Best Local Similarity 100.0%; Pred. No. 4,1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FRNSSIKSYFSDQVSTFRSV 21
87 FRNSSIKSYFSDQVSTFRSV 107
RESULT 10
ADP08984
XX ID ADP08984 standard; peptide; 233 AA.
XX AC ADP08984;
XX DT 12-FEB-2004 (first entry)
```

```
XX DE Secreted ovarian carcinoma antigen segid 488.
XX KW gene therapy; protein therapy; vaccine; antibody inhibition;
XX KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
XX KW secreted ovarian carcinoma antigen.
XX OS Homo sapiens.
XX PN US2003124140-A1.
XX PD 03-JUL-2003.
XX PF 17-JUL-2002; 2002US-00198053.
XX PR 17-DEC-1998; 98US-00215681.
XX PR 17-DEC-1998; 98US-00216003.
XX PR 23-JUN-1999; 99US-00338933.
XX PR 24-SEP-1999; 99US-00404879.
XX PR 17-JUL-2000; 2000US-00617747.
XX PR 10-AUG-2000; 2000US-00636801.
XX PR 20-SEP-2000; 2000US-00667857.
XX PR 04-APR-2001; 2001US-00827271.
XX PR 18-JUN-2001; 2001US-00884441.
XX PR 17-JUL-2001; 2001US-00907969.
XX PA (CORI-) CORIXA CORP.
XX PI Bangur CS, Retter MW, Fanger GR, Hill P;
XX PI WPI; 2003-897152/82.
XX DR Oncogenic nucleic acids useful for the prevention, diagnosis and
XX PT treatment of breast cancer.
XX PS Example 13; SEQ ID NO 488; 399pp; English.
XX CC The invention describes nucleic acids (I) and the polypeptides (II) they
XX CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
XX CC treating diseases related to their aberrant expression i.e. breast
XX CC cancers. For example, (I) and (II) may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of (II) by expressing
XX CC inactive proteins or to supplement the patient's own production of (II).
XX CC Additionally, (I) may be used to produce (II), by inserting (I) into a
XX CC host cell and culturing the cell to express the protein (II). (I) And its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar nucleic acids in
XX CC samples, and therefore which patients may be in need of restorative
XX CC therapy. The host cell may also be used as antigens in the production of
XX CC antibodies against (II) and in assays to identify modulators of (II)'s
XX CC expression and activity. The anti-(II) antibodies, agonists and as
XX CC antagonists may be used to regulate expression and activity and as
XX CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
XX CC immunoassay). This sequence represents a secreted ovarian carcinoma
XX CC antigen.
XX SQ Sequence 233 AA;
OY Query Match 100.0%; Score 107; DB 7; Length 233;
Best Local Similarity 100.0%; Pred. No. 4,1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 FRNSSIKSYFSDQVSTFRSV 21
87 FRNSSIKSYFSDQVSTFRSV 107
RESULT 11
ABU54859
XX ID ABU54859 standard; protein; 284 AA.
XX AC ABU54859;
```

XX 12-MAR-2003 (first entry)
 XX Human CA125 carboxy terminal domain.
 DE Human CA125; protein repeat; chromosome 19q13.2; amino terminal domain;
 XX amino terminal extension; carboxy terminal domain; vaccine; cancer;
 KW ovarian cancer; carcinoma.
 XX Homo sapiens.
 XX WO200283866-A2.
 XX 24-OCT-2002.
 XX 12-APR-2002; 2002WO-US011734.
 XX 17-APR-2001; 2001US-0284175P.
 PR 19-JUN-2001; 2001US-0299380P.
 PR 27-SEP-2001; 2001US-00965738.
 PR 21-DEC-2001; 2001US-0345180P.
 XX (UVAR-) UNIV ARKANSAS.
 PA O'Brien T, Beard J, Underwood L;
 DR WPI; 2003-093013/08.
 XX New CA125 molecules, useful as a gold standard for detecting and
 PT monitoring the presence of CA125 antigen which can be used for
 PT diagnosing, monitoring or treating patients with cancer or for developing
 PT vaccine against cancer.
 XX Claim 1; Fig 9b; 694pp; English.

XX The invention relates to a CA125 protein comprising: (a) an extracellular
 CC amino terminal domain; (b) an amino terminal extension; (c) a multiple
 CC repeat domain; and (d) a carboxy terminal domain. The extracellular amino
 CC terminal domain comprises 5 genomic exons, the amino terminal extension
 CC comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and
 CC the carboxy terminal domain comprises a transmembrane anchor with a short
 CC cytoplasmic domain, and further comprises 9 genomic exons. The gene for
 CC CA125 is located on human chromosome 19q13.2. Also included are isolated
 CC CA125 repeat domains, nucleic acids (including variants, homologues and
 CC degenerate versions) encoding CA125 proteins or repeat units, a vector
 CC comprising the nucleic acid, a cultured cell comprising the vector, a
 CC method of expressing CA125 antigen in a cell, the amino acid sequences of
 CC the CA125 repeat units (or their variants, fragments or sequences 50%
 CC identical to them), a purified antibody that selectively binds to an
 CC epitope in the receptor-binding domain of CA125 protein, a diagnostic for
 CC detecting and monitoring the presence of CA125 antigen (comprising
 CC recombinant CA125 having at least one repeat unit of the CA125 repeat
 CC domain including epitope binding sites), a therapeutic vaccine to treat
 CC mammals with elevated CA125 antigen levels or at risk of developing a
 CC disease or disease recurrence associated with elevated CA125 antigen
 CC levels (comprising recombinant CA125 repeat domain including epitope
 CC binding sites) and an antisense oligonucleotide that inhibits the
 CC expression of CA125. The CA125 molecule, particularly the multiple repeat
 CC domain are useful as a gold standard for detecting and monitoring the
 CC presence of CA125 antigen, which can be used for diagnosing, monitoring
 CC or treating patients with ovarian cancer and other carcinomas where CA125
 CC is expressed. The molecules are also useful for developing a vaccine
 CC against cancer. The present sequence is a CA125 repeat protein

XX Sequence 284 AA;

Query Match 100.0%; Score 107; DB 6; Length 284;
 Best Local Similarity 100.0%; Pred. No. 5.1e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
 DB 138 FRNSIKSYFSDQVSTFRSV 158

RESULT 12
 ID ABP31025 standard; protein; 318 AA.
 XX AC ABP31025;
 XX 02-JUL-2002 (first entry)
 DE Amino acid sequence of the 3' constant region of 0772P.
 XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
 XX Homo sapiens.
 XX WO200206317-A2.
 XX 24-JAN-2002.
 XX 17-JUL-2001; 2001WO-US022635.
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 XX (CORI-) CORIXA CORP.
 PA Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
 PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
 DR WPI; 2002-164781/21.
 XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
 PT protein or its variants, useful for stimulating an immune response in a
 PT patient and treating ovarian cancer.
 XX Claim 3; Page 399-400; 408pp; English.

XX This invention relates to polypeptides comprising an immunogenic portion
 CC of an ovarian carcinoma protein which acts as an immunostimulant and is
 CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations and antigen presenting cells that express
 CC the polypeptides are useful for stimulating an immune response in a
 CC patient and treating ovarian cancer. This sequence represents protein
 CC related to the invention

XX Sequence 318 AA;

Query Match 100.0%; Score 107; DB 5; Length 318;
 Best Local Similarity 100.0%; Pred. No. 5.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
 DB 172 FRNSIKSYFSDQVSTFRSV 192

RESULT 13
 ID ADA08747 standard; protein; 318 AA.
 XX AC ADA08747;
 XX 06-NOV-2003 (first entry)

Query Match 100.0%; Score 107; DB 5; Length 318;
 Best Local Similarity 100.0%; Pred. No. 5.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
 DB 172 FRNSIKSYFSDQVSTFRSV 192


```

XX XX US2003091580-A1.
XX PN
XX 15-MAY-2003.
XX PD
XX 17-JUL-2001; 2001US-00907969.
XX PF
XX 18-JUN-2001; 2001US-00844441.
XX PR
XX (MITC/) MITCHAM J L.
XX (KING/) KING G E.
XX (ALGA/) ALGATE P A.
XX (FLIN/) FLING S P.
XX (RETT/) RETTER M W.
XX (FANG/) FANGER G R.
XX (REED/) REED S G.
XX (VEDV/) VEDVICK T S.
XX (CART/) CARTER D.
XX (HILL/) HILL P.
XX (ALBO/) ALBONE E.
XX
XX MITCHAM JL, KING GE, Algate PA, Fling SP, Retter MW, Fanger GR;
XX PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX DR WPI; 2003-532352/50.
XX
XX New isolated 0772P polypeptides and polynucleotides, useful in gene
XX PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
XX cancer.
XX
XX Claim 3; SEQ ID NO 594; 371pp; English.
XX
XX The invention relates to an isolated 0772P polypeptide, which has the
XX CC structure fully defined in the specification. The composition containing
XX CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
XX CC or antigen presenting cells are useful for stimulating an immune response
XX CC and treating ovarian cancer. Detecting the presence of the
XX CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
XX CC carcinoma cDNAs and protein cDNAs were identified using microarray
XX CC technology. The present sequence represents a human ovarian carcinoma
XX CC antigen.
XX
XX SQ Sequence 318 AA;

```

```

Query Match          100.0%; Score 107; DB 7; Length 318;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 FRNSSTKSYFSDCVSTFRSV 21
   |||||
Db 172 FRNSSTKSYFSDCVSTFRSV 192

```

```

RESULT 14
ADF09090
ID ADF09090 standard; protein; 318 AA.
AC ADF09090;
XX
XX 12-FEB-2004 (first entry)
XX
XX Secreted ovarian carcinoma antigen seqid 594.
XX
XX gene therapy; protein therapy; vaccine; antibody inhibition;
XX KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
XX KW secreted ovarian carcinoma antigen.
XX
XX Homo sapiens.
XX OS
XX US2003124140-A1.
XX PN
XX 03-JUL-2003.
XX PD
XX

```

```

PF 17-JUL-2002; 2002US-00196053.
XX
XX 17-DEC-1998; 98US-00215681.
XX PR 17-DEC-1998; 98US-00216003.
XX PR 23-JUN-1999; 99US-00338933.
XX PR 24-SEP-1999; 99US-00404879.
XX PR 17-JUL-2000; 2000US-00617747.
XX PR 10-AUG-2000; 2000US-00636801.
XX PR 20-SEP-2000; 2000US-00667857.
XX PR 04-APR-2001; 2001US-00827271.
XX PR 18-JUN-2001; 2001US-00844441.
XX PR 17-JUL-2001; 2001US-00907969.
XX
XX (CORI-) CORIXA CORP.
XX
XX Bangur CS, Retter MW, Fanger GR, Hill P;
XX PI WPI; 2003-697152/82.
XX DR N-PSDB; ADF09064.
XX
XX Oncogenic nucleic acids useful for the prevention, diagnosis and
XX PT treatment of breast cancer.
XX
XX Example 16; SEQ ID NO 594; 399pp; English.
XX
XX The invention describes nucleic acids (I) and the polypeptides (II) they
XX CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
XX CC treating diseases related to their aberrant expression i.e. breast
XX CC cancers. For example, (I) and (II) may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of (II) by expressing
XX CC inactive proteins or to supplement the patient's own production of (II).
XX CC Additionally, (II) may be used to produce (II), by inserting (I) into a
XX CC host cell and culturing the cell to express the protein (II). (I) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar nucleic acids in
XX CC samples, and therefore which patients may be in need of restorative
XX CC therapy. The host cell may also be used as antigens in the production of
XX CC antibodies against (II) and in assays to identify modulators of (II)'s
XX CC expression and activity. The anti-(II) antibodies, agonists and
XX CC antagonists may be used to regulate expression and activity and as
XX CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
XX CC immunoassay). This sequence represents a secreted ovarian carcinoma
XX CC antigen.
XX
XX SQ Sequence 318 AA;

```

```

Query Match          100.0%; Score 107; DB 7; Length 318;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 FRNSSTKSYFSDCVSTFRSV 21
   |||||
Db 172 FRNSSTKSYFSDCVSTFRSV 192

```

```

RESULT 15
AAE12631
ID AAE12631 standard; protein; 367 AA.
AC AAE12631;
XX
XX 03-JUN-2002 (first entry)
XX
XX Human gene 1 encoded secreted protein fragment, SEQ ID NO: 27.
XX
XX Human; ovarian cancer antigen; proliferative disorder; cancer; tumour;
XX KW respiratory system disorder; asthma; haematopoietic disorder; skin aging;
XX KW immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
XX KW rheumatoid arthritis; inflammation; neurological disorder; septic shock;
XX KW Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
XX KW atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
XX KW epithelial cell proliferation; transplantation; chemotaxis; infection;

```

KW food additive; wound healing; endocrine disorder; kidney disorder;
KW gene therapy; cytostatic.
XX
XX Homo sapiens.
OS
PN WO200170804-A1.
XX
PD 27-SEP-2001.
XX
PF 16-MAR-2001; 2001WO-US008585.
XX
PR 17-MAR-2000; 2000US-0190076P.
PR 23-AUG-2000; 2000US-0227009P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2001-639119/73.
XX
XX Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
PT metastases.
XX
PS Disclosure; Page 9; 427pp; English.
XX
CC AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
CC ovarian cancer associated protein (collectively known as ovarian cancer
CC antigens) genes, and AAI12623-AAI12629 represent the proteins they
CC encode. AAI12630-AAI12638 represent human ovarian cancer antigenic
CC fragments or variants. Ovarian cancer antigens and their corresponding
CC DNAs are used in the prevention, diagnosis and treatment of diseases
CC associated with their inappropriate expression. These disorders include
CC proliferative disorders, cancer, tumours, respiratory system disorders,
CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
CC skin disorders, autoimmune diseases (e.g., rheumatoid arthritis),
CC inflammation, allergies, neurological disorders (e.g., Alzheimer's
CC disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, endocrine disorders and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, to identify their cognate ligands or binding
CC partners, in chemotaxis and can be used as a food additive. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above and in diagnostic
CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
CC the invention is used in gene therapy. The present sequence represents a
CC human ovarian cancer antigenic fragment of the invention
XX
SQ Sequence 367 AA;
XX

Query Match 100.0%; Score 107; DB 4; Length 367;
Best Local Similarity 100.0%; Pred. No. 6.8e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSSIKSYFSDQVSTFRSV 21
|||
DB 292 FRNSSIKSYFSDQVSTFRSV 312

Search completed: October 27, 2005, 18:44:24
Job time : 72.6698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:03:42 ; Search time 18.2651 Seconds
(without alignments)
85.826 Million cell updates/sec

Title: US-10-612-090-16
Perfect score: 107
Sequence: 1 FRNSSIKSYFSDQVSTFRSV 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	438	4 US-09-404-879A-390	Sequence 390, App
2	107	100.0	438	4 US-09-667-857-390	Sequence 390, App
3	107	100.0	833	4 US-09-404-879A-389	Sequence 389, App
4	107	100.0	833	4 US-09-667-857-389	Sequence 389, App
5	107	100.0	914	4 US-09-404-879A-312	Sequence 312, App
6	107	100.0	914	4 US-09-338-933-312	Sequence 312, App
7	107	100.0	914	4 US-09-667-857-312	Sequence 312, App
8	46	43.0	144	4 US-09-107-433-3451	Sequence 3451, App
9	46	43.0	509	3 US-09-425-383-2	Sequence 2, App1
10	42	39.3	366	4 US-09-787-069-2	Sequence 1356, App
11	42	39.3	374	4 US-09-710-279-1356	Sequence 5234, App
12	42	39.3	391	3 US-09-134-001C-5234	Sequence 15403, A
13	42	39.3	414	4 US-09-248-796A-15403	Sequence 16, App1
14	42	39.3	561	3 US-09-442-100-16	Sequence 16, App1
15	42	39.3	561	4 US-08-339-106-16	Sequence 16, App1
16	42	39.3	561	4 US-09-442-102-16	Sequence 16, App1
17	42	39.3	772	4 US-09-404-879A-388	Sequence 388, App
18	42	39.3	772	4 US-09-667-857-388	Sequence 388, App
19	40.5	37.9	44	1 US-08-145-708A-1	Sequence 1, App1
20	40.5	37.9	44	2 US-08-331-454-1	Sequence 1, App1
21	40	37.4	91	4 US-09-107-532A-3885	Sequence 3885, App
22	40	37.4	94	4 US-09-543-681A-6936	Sequence 6936, App
23	40	37.4	369	4 US-09-248-796A-18337	Sequence 18337, A
24	40	37.4	587	4 US-09-270-767-37567	Sequence 37567, A
25	40	37.4	587	4 US-09-270-767-37567	Sequence 52784, A
26	40	37.4	683	4 US-09-620-412C-357	Sequence 357, App
27	40	37.4	683	4 US-09-598-419-357	Sequence 357, App

28	40	37.4	948	4 US-09-556-877-194	Sequence 194, App
29	40	37.4	948	4 US-09-620-412C-194	Sequence 194, App
30	40	37.4	948	4 US-09-598-419-194	Sequence 194, App
31	40	37.4	1065	4 US-09-538-092-84	Sequence 84, App1
32	40	37.4	1776	4 US-09-556-877-179	Sequence 179, App
33	40	37.4	1776	4 US-09-620-412C-179	Sequence 179, App
34	40	37.4	1776	4 US-09-598-419-179	Sequence 179, App
35	40	37.4	2476	4 US-09-824-574-7	Sequence 7, App1
36	39.5	36.9	360	4 US-09-538-092-471	Sequence 471, App
37	39	36.4	81	4 US-09-513-999C-7355	Sequence 7355, App
38	39	36.4	97	3 US-08-894-699-27	Sequence 27, App1
39	39	36.4	97	3 US-09-444-410-27	Sequence 27, App1
40	39	36.4	108	4 US-09-543-681A-6306	Sequence 6306, App
41	39	36.4	174	4 US-09-248-796A-18328	Sequence 18328, A
42	39	36.4	248	4 US-09-543-681A-5155	Sequence 5155, App
43	39	36.4	320	2 US-08-530-165-7	Sequence 7, App1
44	39	36.4	375	3 US-09-134-001C-5079	Sequence 5079, App
45	39	36.4	636	3 US-09-142-791A-2	Sequence 2, App1

ALIGNMENTS

```

RESULT 1
; US-09-404-879A-390
; Sequence 390, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-404-879A-390

Query Match      100.0%; Score 107; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 7.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 FRNSSIKSYFSDQVSTFRSV 21
      |||||
Db      292 FRNSSIKSYFSDQVSTFRSV 312

RESULT 2
; US-09-667-857-390
; Sequence 390, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455

```

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 390
LENGTH: 438
TYPE: PRT
ORGANISM: Homo sapiens
US-09-667-857-390

Query Match 100.0%; Score 107; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 7,3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
|||||
Db 292 FRNSIKSYFSDCOVSTFRSV 312

RESULT 3
US-09-404-879A-389
Sequence 389, Application US/09404879A
Patent No. 6468546

GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 389
LENGTH: 833
TYPE: PRT
ORGANISM: Homo sapiens
US-09-404-879A-389

Query Match 100.0%; Score 107; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1,5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
|||||
Db 687 FRNSIKSYFSDCOVSTFRSV 707

RESULT 4
US-09-667-857-389
Sequence 389, Application US/09667857
Patent No. 6699664

GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
APPLICANT: Algate, Paul A.
APPLICANT: Fling, Steven P.
APPLICANT: Retter, Marc W.
APPLICANT: Panger, Gary Richard
APPLICANT: Reed, Steven G.
APPLICANT: Vedvick, Thomas S.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C5
CURRENT APPLICATION NUMBER: US/09/667,857
CURRENT FILING DATE: 2000-09-20
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 389
LENGTH: 833
TYPE: PRT
ORGANISM: Homo sapiens
US-09-667-857-389

Query Match 100.0%; Score 107; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1,5e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
|||||
Db 687 FRNSIKSYFSDCOVSTFRSV 707

RESULT 5
US-09-404-879A-312
Sequence 312, Application US/09404879A
Patent No. 6468546

GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 312
LENGTH: 914
TYPE: PRT
ORGANISM: Homo sapien
US-09-404-879A-312

Query Match 100.0%; Score 107; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1,7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
|||||
Db 768 FRNSIKSYFSDCOVSTFRSV 788

RESULT 6
US-09-338-933-312
Sequence 312, Application US/09338933
Patent No. 6488931

GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer Lynn
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
FILE REFERENCE: 210121.462C1
CURRENT APPLICATION NUMBER: US/09/338,933
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 312
LENGTH: 914
TYPE: PRT
ORGANISM: Homo sapien
US-09-338-933-312

Query Match 100.0%; Score 107; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1,7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
|||||
Db 768 FRNSIKSYFSDCOVSTFRSV 788

RESULT 7
US-09-667-857-312

Sequence 312, Application US/09667857
Patent No. 6699664
GENERAL INFORMATION:

```
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: King, Gordon B.
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Rling, Steven P.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Fanger, Gary Richard
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.462C5
/ CURRENT APPLICATION NUMBER: US/09/667,857
/ CURRENT FILING DATE: 2000-09-20
/ NUMBER OF SEQ ID NOS: 455
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 312
/ LENGTH: 914
/ TYPE: PRT
/ ORGANISM: Homo sapien
/ US-09-667-857-312

Query Match          100.0%; Score 107; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
DB 768 FRNSIKSYFSDQVSTFRSV 788

RESULT 8
US-09-107-433-3451
/ Sequence 3451, Application US/09107433
/ Patent No. 6800744
/ GENERAL INFORMATION:
/ APPLICANT: Lynn A Doucette-Stamm and David Bush
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
/ SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
/ THERAPEUTICS
/ NUMBER OF SEQUENCES: 5206
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: GENOME THERAPEUTICS CORPORATION
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD-ROM ISO9660
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneka
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-8277
/ TELEFAX: (781)893-8277
/ INFORMATION FOR SEQ ID NO: 3451:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 144 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
```

```
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (8) LOCATION 1...144
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3451:
/ US-09-107-433-3451

Query Match          43.0%; Score 46; DB 4; Length 144;
Best Local Similarity 56.2%; Pred. No. 3.8;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 NSSIKSYFSDQVSTF 18
DB 55 NSSPVSVMRCSLSLF 70

RESULT 9
US-09-425-383-2
/ Sequence 2, Application US/09425383
/ Patent No. 6194637
/ GENERAL INFORMATION:
/ APPLICANT: Mahajan, Pramod B.
/ APPLICANT: Shi, Jinru
/ TITLE OF INVENTION: Maize DNA Ligase I Orthologue and Uses
/ TITLE OF INVENTION: Thereof
/ FILE REFERENCE: 0962
/ CURRENT APPLICATION NUMBER: US/09/425,383
/ CURRENT FILING DATE: 1999-10-22
/ EARLIER APPLICATION NUMBER: 60/108,793
/ EARLIER FILING DATE: 1998-11-17
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 909
/ TYPE: PRT
/ ORGANISM: Zea mays
/ US-09-425-383-2

Query Match          43.0%; Score 46; DB 3; Length 909;
Best Local Similarity 33.3%; Pred. No. 32;
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTF 18
DB 603 FRKATVKSFLDCEIVAY 620

RESULT 10
US-09-787-069-2
/ Sequence 2, Application US/09787069
/ Patent No. 6627429
/ GENERAL INFORMATION:
/ APPLICANT: Danisco A/S
/ APPLICANT: Christensen, Tove MIE
/ APPLICANT: Pedersen, Anette A
/ APPLICANT: Brunstedt, Janne
/ APPLICANT: Mikkelsen, Jorn D
/ TITLE OF INVENTION: Process
/ FILE REFERENCE: P005380MO.CTH
/ CURRENT APPLICATION NUMBER: US/09/787,069
/ CURRENT FILING DATE: 2001-07-16
/ PRIOR APPLICATION NUMBER: GB 9820195.7
/ PRIOR FILING DATE: 1998-09-16
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 366
/ TYPE: PRT
/ ORGANISM: Erwinia chrysanthemi
/ US-09-787-069-2
```

Query Match 39.3%; Score 42; DB 4; Length 366;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 SIKSYFSDCCQVSTR 16
DB 186 RSFSDCRIS 195

RESULT 11
US-09-710-279-1356
; Sequence 1356, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUJ480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1356
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-1356

Query Match 39.3%; Score 42; DB 4; Length 374;
Best Local Similarity 37.5%; Pred. No. 54;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 4 SIKSYFSDCCQVSTR 19
DB 132 SEVQNYFKDKQIETVQ 147

RESULT 12
US-09-134-001C-5234
; Sequence 5234, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5234
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5234

Query Match 39.3%; Score 42; DB 3; Length 391;
Best Local Similarity 37.5%; Pred. No. 57;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 4 SIKSYFSDCCQVSTR 19
DB 149 SEVQNYFKDKQIETVQ 164

RESULT 13
US-09-248-796A-15403
; Sequence 15403, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15403
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15403

Query Match 39.3%; Score 42; DB 4; Length 414;
Best Local Similarity 44.4%; Pred. No. 61;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 SIKSYFSDCCQVSTRSV 21
DB 281 AEVQYFSDALSDGKSV 298

RESULT 14
US-09-442-100-16
; Sequence 16, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Weiyl
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,100
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/411,111
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mastrock, S. Leelle
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

LENGTH: 561 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-09-442-100-16

Query Match 39.3%; Score 42; DB 3; length 561;
 Best Local Similarity 42.9%; Pred. No. 87;
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 FRNSIKSYFSDQVSTFRSV 21
 DB 457 FEHVKMSYFADINFSTLRSM 477

RESULT 15

US-08-939-106-16
 Sequence 16, Application US/08939106
 Patent No. 6559285
 GENERAL INFORMATION:

APPLICANT: Yale University
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
 GENES AND METHODS BASED THEREON
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/939,106
 FILING DATE: 26-No. 6559285-1997
 CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Mastrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 6523-007-228
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 561 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown

MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 16:
 US-08-939-106-16

Query Match 39.3%; Score 42; DB 4; length 561;
 Best Local Similarity 42.9%; Pred. No. 87;
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 FRNSIKSYFSDQVSTFRSV 21
 DB 457 FEHVKMSYFADINFSTLRSM 477

Search completed: October 27, 2005, 17:34:33
 Job time : 19.2651 secs

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:30:08 ; Search time 67.2 Seconds
(without alignments)
130.638 Million cell updates/sec

Title: US-10-612-090-16
Perfect score: 107
Sequence: 1 FRNSSIKSYFSDCOVSTFRSV 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1865214 seqs, 418043040 residues

Total number of hits satisfying chosen parameters: 1865214

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	21	US-10-612-090-16	Sequence 16, App1
2	107	100.0	108	US-10-612-090-2	Sequence 2, App1
3	107	100.0	108	US-10-612-090-13	Sequence 13, App1
4	107	100.0	174	US-10-383-368-4	Sequence 4, App1
5	107	100.0	178	US-09-884-441-489	Sequence 489, App
6	107	100.0	178	US-09-907-969-489	Sequence 489, App
7	107	100.0	178	US-10-198-053-489	Sequence 489, App
8	107	100.0	178	US-10-860-790-489	Sequence 489, App
9	107	100.0	233	US-09-884-441-488	Sequence 488, App
10	107	100.0	233	US-09-907-969-488	Sequence 488, App
11	107	100.0	233	US-10-198-053-488	Sequence 488, App

12	107	100.0	233	17	US-10-860-790-488	Sequence 488, App
13	107	100.0	284	10	US-09-965-738-300	Sequence 300, App
14	107	100.0	318	10	US-09-907-969-594	Sequence 594, App
15	107	100.0	318	14	US-10-198-053-594	Sequence 594, App
16	107	100.0	318	14	US-10-860-790-594	Sequence 594, App
17	107	100.0	367	15	US-10-333-900-27	Sequence 27, App1
18	107	100.0	396	18	US-10-858-412-225	Sequence 225, App
19	107	100.0	438	9	US-09-884-441-390	Sequence 390, App
20	107	100.0	438	9	US-09-884-441-483	Sequence 483, App
21	107	100.0	438	10	US-09-907-969-390	Sequence 390, App
22	107	100.0	438	10	US-09-907-969-483	Sequence 483, App
23	107	100.0	438	10	US-09-827-271-390	Sequence 390, App
24	107	100.0	438	14	US-10-198-053-390	Sequence 390, App
25	107	100.0	438	14	US-10-198-053-483	Sequence 483, App
26	107	100.0	438	15	US-10-333-900-19	Sequence 19, App1
27	107	100.0	438	17	US-10-860-790-390	Sequence 390, App
28	107	100.0	438	17	US-10-860-790-483	Sequence 483, App
29	107	100.0	439	10	US-09-965-738-148	Sequence 148, App
30	107	100.0	456	18	US-10-858-412-226	Sequence 226, App
31	107	100.0	526	15	US-10-333-900-30	Sequence 30, App1
32	107	100.0	545	14	US-10-243-243A-4	Sequence 4, App1
33	107	100.0	583	14	US-10-142-515-4	Sequence 4, App1
34	107	100.0	748	17	US-10-687-035-1	Sequence 1, App1
35	107	100.0	809	17	US-10-687-035-2	Sequence 2, App1
36	107	100.0	833	9	US-09-884-441-389	Sequence 389, App
37	107	100.0	833	10	US-09-907-969-389	Sequence 389, App
38	107	100.0	833	10	US-09-827-271-389	Sequence 389, App
39	107	100.0	833	14	US-10-198-053-389	Sequence 389, App
40	107	100.0	833	17	US-10-860-790-389	Sequence 389, App
41	107	100.0	914	9	US-09-178-320-206	Sequence 206, App
42	107	100.0	914	9	US-09-910-689-206	Sequence 206, App
43	107	100.0	914	9	US-09-884-441-312	Sequence 312, App
44	107	100.0	914	9	US-09-884-441-478	Sequence 478, App
45	107	100.0	914	10	US-09-907-969-312	Sequence 312, App

ALIGNMENTS

RESULT 1
US-10-612-090-16
; Sequence 16, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-612-090-16

Query Match 100.0%, Score 107; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 FRNSSIKSYFSDCOVSTFRSV 21
|||
Db 1 FRNSSIKSYFSDCOVSTFRSV 21

RESULT 2
US-10-612-090-2
; Sequence 2, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:

```
/ APPLICANT: ImmunoGen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-612-090-2

Query Match          100.0%; Score 107; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 7,3e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
   |||||
Db 18 FRNSIKSYFSDCOVSTFRSV 38

RESULT 3
US-10-612-090-13
/ Sequence 13, Application US/10612090
/ Publication No. US20040057952A1
/ GENERAL INFORMATION:
/ APPLICANT: ImmunoGen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 13
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURES:
/ OTHER INFORMATION: Fusion protein
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (1)-(1)
/ OTHER INFORMATION: Glutathione S-transferase fusion site
US-10-612-090-13

Query Match          100.0%; Score 107; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 7,3e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
   |||||
Db 18 FRNSIKSYFSDCOVSTFRSV 38

RESULT 4
US-10-383-368-4
/ Sequence 4, Application US/10383368
/ Publication No. US20040002091A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul E.
/ APPLICANT: Ebner, Reinhard
/ APPLICANT: Strovel, Jeffrey W.
/ TITLE OF INVENTION: Cancer-Linked Gene as Target for Chemotherapy
/ FILE REFERENCE: 689290-123
/ CURRENT APPLICATION NUMBER: US/10/383,368
/ CURRENT FILING DATE: 2003-03-07
/ PRIOR APPLICATION NUMBER: US 60/362,527
/ PRIOR FILING DATE: 2002-03-07
/ ORGANISM: Homo sapiens
```

```
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 4
/ LENGTH: 174
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-383-368-4

Query Match          100.0%; Score 107; DB 15; Length 174;
Best Local Similarity 100.0%; Pred. No. 1,2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
   |||||
Db 28 FRNSIKSYFSDCOVSTFRSV 48

RESULT 5
US-09-884-441-489
/ Sequence 489, Application US/09884441
/ Patent No. US20020119158A1
/ GENERAL INFORMATION:
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Carter, Derrick
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.462C7
/ CURRENT APPLICATION NUMBER: US/09/884,441
/ CURRENT FILING DATE: 2001-06-18
/ NUMBER OF SEQ ID NOS: 489
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 489
/ LENGTH: 178
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-884-441-489

Query Match          100.0%; Score 107; DB 9; Length 178;
Best Local Similarity 100.0%; Pred. No. 1,2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
   |||||
Db 87 FRNSIKSYFSDCOVSTFRSV 107

RESULT 6
US-09-907-969-489
/ Sequence 489, Application US/09907969
/ Publication No. US20030091580A1
/ GENERAL INFORMATION:
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: King, Gordon E.
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Fling, Steven P.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Fanger, Gary Richard
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Vedick, Thomas S.
/ APPLICANT: Carter, Derrick
/ APPLICANT: Hill, Paul
/ APPLICANT: Albone, Earl
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.462C8
/ CURRENT APPLICATION NUMBER: US/09/907,969
/ CURRENT FILING DATE: 2001-07-17
/ NUMBER OF SEQ ID NOS: 596
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 489
/ LENGTH: 178
/ TYPE: PRT
/ ORGANISM: Homo sapiens
```

US-09-907-969-489

Query Match 100.0%; Score 107; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
Db 87 FRNSIKSYFSDQVSTFRSV 107

RESULT 7

US-10-198-053-489
; Sequence 489, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-489

Query Match 100.0%; Score 107; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
Db 87 FRNSIKSYFSDQVSTFRSV 107

RESULT 8
US-10-860-790-489
; Sequence 489, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-489

Query Match 100.0%; Score 107; DB 17; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
Db 87 FRNSIKSYFSDQVSTFRSV 107

RESULT 9

US-09-884-441-488
; Sequence 488, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-488

Query Match 100.0%; Score 107; DB 9; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
Db 87 FRNSIKSYFSDQVSTFRSV 107

RESULT 10
US-09-907-969-488
; Sequence 488, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-969-488

Query Match 100.0%; Score 107; DB 10; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDQVSTFRSV 21
Db 87 FRNSIKSYFSDQVSTFRSV 107

RESULT 11
US-10-198-053-488
; Sequence 488, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.

```

APPLICANT: Reltter, Marc W.
APPLICANT: Fanger, Gary R.
APPLICANT: Hill, Paul
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C9
CURRENT APPLICATION NUMBER: US/10/198.053
CURRENT FILING DATE: 2002-07-17
NUMBER OF SEQ ID NOS: 624
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 488
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
US-10-198-053-488

Query Match          100.0%; Score 107; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FRNSSIKSYFSDCCVSTFRSV 21
      |||||
Db      87 FRNSSIKSYFSDCCVSTFRSV 107

RESULT 12
US-10-860-790-488
Sequence 488, Application US/10860790
Publication No. US20050031634A1
GENERAL INFORMATION:
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Reltter, Marc W.
APPLICANT: Fanger, Gary R.
APPLICANT: Hill, Paul
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C11
CURRENT APPLICATION NUMBER: US/10/860.790
CURRENT FILING DATE: 2004-06-02
NUMBER OF SEQ ID NOS: 624
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 488
LENGTH: 233
TYPE: PRT
ORGANISM: Homo sapiens
US-10-860-790-488

Query Match          100.0%; Score 107; DB 17; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 FRNSSIKSYFSDCCVSTFRSV 21
      |||||
Db      87 FRNSSIKSYFSDCCVSTFRSV 107

RESULT 13
US-09-965-738-300
Sequence 300, Application US/09965738
Publication No. US20030143667A1
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy
TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic
TITLE OF INVENTION: Therapeutic Interventions
FILE REFERENCE: 40715-258841
CURRENT APPLICATION NUMBER: US/09/965.738
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US 60/284,175
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 306
SOFTWARE: PatentIn version 3.0
SEQ ID NO 300
LENGTH: 284

```

```

: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-965-738-300

Query Match          100.0%; Score 107; DB 10; Length 284;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 FRNSSIKSYFSDCOVSTFRSV 21
      |||||||||||||||||||
Db      138 FRNSSIKSYFSDCOVSTFRSV 158

RESULT 14
US-09-907-969-594
; Sequence 594, Application US/09907969
; Publication No. US20030091580A1
GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Ranger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C8
CURRENT APPLICATION NUMBER: US/09/907,969
CURRENT FILING DATE: 2001-07-17
NUMBER OF SEQ ID NOS: 596
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 594
LENGTH: 318
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Variant
LOCATION: 136,248,268
OTHER INFORMATION: Xaa = Any amino acid
US-09-907-969-594

Query Match          100.0%; Score 107; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 FRNSSIKSYFSDCOVSTFRSV 21
      |||||||||||||||||||
Db      172 FRNSSIKSYFSDCOVSTFRSV 192

RESULT 15
US-10-198-053-594
; Sequence 594, Application US/10198053
; Publication No. US20030124140A1
GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Ranger, Gary R.
; APPLICANT: Hill, Paul
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C9
CURRENT APPLICATION NUMBER: US/10/198,053
CURRENT FILING DATE: 2002-07-17
NUMBER OF SEQ ID NOS: 624
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 594
LENGTH: 318

```

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: variant
LOCATION: 136,248,268
OTHER INFORMATION: Xaa = Any amino acid
US-10-198-053-594

Query Match 100.0%; Score 107; DB 14; Length 318;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FRNSSIKSYFSDQVSTFRSV 21
Db 172 FRNSSIKSYFSDQVSTFRSV 192

Search completed: October 27, 2005, 18:31:46
Job time : 67.2 secs

Robert Young Campbell

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 18:13:49 ; Search time 12.9907 Seconds
(without alignments)
155.538 Million cell updates/sec

Title: US-10-612-090-16
Perfect score: 107
Sequence: 1 FRNSSIKSYFSDQVSTFRSV 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	50.5	496	2 E71192	probable TRK sysce
2	49	45.8	493	2 G75151	trk potassium upta
3	48	44.9	809	2 T40574	guanine nucleotide
4	47	43.9	902	2 S33918	dynammin-like prote
5	44	41.1	51	2 S42357	probable zinc fing
6	44	40.2	145	2 C70113	probable transcrip
7	43	40.2	361	2 AG0052	probable pectinest
8	43	40.2	401	2 F97260	uncharacterized co
9	43	40.2	790	2 S71278	DNA ligase (ATP) (
10	43	40.2	1148	2 T00016	minor outer capsid
11	42	39.3	366	2 S03770	pectinesterase (BC
12	42	39.3	366	2 JN0799	pectinesterase (BC
13	42	39.3	368	2 B87753	protein C43E11.5 (
14	42	39.3	368	2 T29779	hypothetical prote
15	42	39.3	494	2 AB0252	glucose-6-phosphat
16	42	39.3	572	2 S64387	protein kinase DBF
17	41.5	38.8	298	2 F70140	glycerol-3-phospha
18	41	38.3	179	2 F75392	hypothetical prote
19	41	38.3	414	2 C86342	F9H16.7 protein-
20	41	38.3	903	2 T20804	hypothetical prote
21	41	38.3	972	2 F71608	hypothetical prote
22	41	38.3	1520	2 B82274	conserved hypotet
23	41	38.3	1612	2 JCS210	DNA (cytosine-5-)-
24	41	38.3	2632	2 T18718	dynamin heavy chain
25	40.5	37.9	93	2 T51188	small zinc finger-
26	40.5	37.9	310	2 S20889	superantigen Mtv1
27	40.5	37.9	312	2 S34635	superantigen Mtv1
28	40.5	37.9	315	2 JH0554	superantigen Mtv13
29	40.5	37.9	315	2 JH0551	superantigen Mtv1/

30	40.5	37.9	320	2 S26174	superantigen Mtv17
31	40.5	37.9	322	2 S24574	superantigen Mtv15
32	40.5	37.9	2014	2 S46622	probable membrane
33	40	37.4	115	2 T39139	hypothetical prote
34	40	37.4	172	2 AD0062	conserved hypotet
35	40	37.4	236	2 S55297	hypothetical prote
36	40	37.4	342	2 E97522	sensory rhodopsin
37	40	37.4	342	2 E97522	peptide chain rele
38	40	37.4	362	2 AF2741	hypothetical prote
39	40	37.4	382	2 A38545	hypothetical prote
40	40	37.4	382	2 E97742	hypothetical prote
41	40	37.4	429	2 T29711	hypothetical prote
42	40	37.4	433	2 G90083	hypothetical prote
43	40	37.4	444	2 T15907	hypothetical prote
44	40	37.4	468	2 S47447	hypothetical prote
45	40	37.4	491	1 D64947	glucose-6-phosphat

ALIGNMENTS

```
RESULT 1
E71192
probable TRK system potassium uptake protein - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: E71192
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71192
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-496 <KAW>
A:Cross-references: UNIPROT:O59477; GB:AP000007; NID:g3236134; PIDN:BAA30932.1; PID:g325f
A:Experimental source: strain O73
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1813

Query Match          50.5%; Score 54; DB 2; Length 496;
Best Local Similarity 58.8%; Pred. No. 0.82;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 FRNSSIKSYFSDQVST 17
      ||| ||| ||| |||
Db      270 FKNRSIRSPFMDIQVT 286

RESULT 2
G75151
trk potassium uptake system protein (trkH) PAB0248 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: G75151
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A:Reference number: A75001
A:Accession: G75151
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <KAW>
A:Cross-references: UNIPROT:Q9V1Q5; GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49294
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: trkH; PAB0248
C:Superfamily: potassium uptake protein trkG

Query Match          45.8%; Score 49; DB 2; Length 493;
Best Local Similarity 47.1%; Pred. No. 5.2;
```

Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVST 17
 ||:|||||
 DB 266 FRDKSLRNFPNDIOVKT 282

RESULT 3
 T40574
 guanine nucleotide binding protein beta subunit-like - fission yeast (Schizosaccharomyces
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T40574
 R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: 221938
 A:Accession: T40574
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-809 <SEB>
 A:Cross-references: UNIPROT:Q94527; EMBL:AL035226; PIDN:CAA22832.1; GSPDB:GN00067
 A:Experimental source: strain 972h-; cosmid c609
 C:Genetics:
 A:Gene: SPAC609.03
 A:Map position: 2
 A:introns: 21/1; 52/2; 280/3; 780/2

Query Match 44.9%; Score 48; DB 2; Length 809;
 Best Local Similarity 36.8%; Pred. No. 12;
 Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 3 NSIKSYFSDCOVSTFRSY 21
 ||:|||||
 DB 658 NTHVKSYYGCHVESIKNV 676

RESULT 4
 S33918
 dynamn-like protein MGMT - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein O5010; protein YOR211c; protein YOR50-1
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C:Accession: S33918; S36728; A42137; S48883; S6038; S67104; S67103; S71713
 R:Guan, K.; Fath, U.; Marshall, T.K.; Deschenes, R.U.
 Curr. Genet. 24, 141-148, 1993
 A:Title: Normal mitochondrial structure and genome maintenance in yeast requires the dyn
 A:Reference number: S33918; MUID:93365024; PMID:7916673
 A:Accession: S33918
 A:Molecule type: DNA
 A:Residues: 1-902 <GUA1>
 A:Cross-references: UNIPROT:P32266; EMBL:L07419
 A>Note: It is uncertain whether Met-1 or Met-60 is the initiator
 R:Guan, K.; Fath, U.; Marshall, T.; Deschenes, R.U.
 submitted to the EMBL Data Library, January 1992
 A:Reference number: S36728
 A:Accession: S36728
 A:Molecule type: DNA
 A:Residues: 1-190, 'A', 192-902 <GUA2>
 A:Cross-references: EMBL:L07419
 R:Jones, B.A.; Pandman, W.L.
 Genes Dev. 6, 380-389, 1992
 A:Title: Mitochondrial DNA maintenance in yeast requires a protein containing a region x
 A:Reference number: A42137; MUID:92192451; PMID:1532158
 A:Accession: A42137
 A:Molecule type: DNA
 A:Residues: 60-149, 'C', 151-902 <JON>
 A:Cross-references: EMBL:X62834; NID:g93956; PIDN:CAA4637.1; PID:g3957
 A>Note: Sequence extracted from NCBI backbone (NCBIN:88065, NCBI:P:88066)
 R:Laio, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
 Proc. Natl. Acad. Sci. U.S.A. 90, 5524-5528, 1993
 A:Title: Interactions between three common subunits of yeast RNA polymerases I and III.
 A:Reference number: S48883; MUID:93396170; PMID:8516295
 A:Accession: S48883

A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 734-902 <IAL>
 A:Cross-references: EMBL:L11274; NID:g295632; PIDN:AAB59316.1; PID:g295633
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1993
 R:Galison, F.; Dujon, B.
 submitted to the EMBL Data Library, October 1995
 A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome
 A:Reference number: S60938
 A:Accession: S60938
 A:Molecule type: DNA
 A:Residues: 1-805 <GAL>
 A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63174.1; PID:g1050763
 R:Boyer, J.; Fairhead, C.; Gallion, L.; Galison, F.; Michaux, G.; Thierry, A.; Dujon, B.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67104
 A:Accession: S67104
 A:Molecule type: DNA
 A:Residues: 1-805 <BOY>
 A:Cross-references: EMBL:Z75119; MIPS:YOR211c
 R:Hughes, B.; Pohl, T.M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66685
 A:Accession: S67103
 A:Molecule type: DNA
 A:Residues: 179-902 <HUG>
 A:Cross-references: EMBL:Z75119; MIPS:YOR211c
 A:Experimental source: strain S288C
 R:Galison, F.; Dujon, B.
 yeast 12, 877-885, 1996
 A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV of
 A:Reference number: S71713; MUID:96437977; PMID:8840505
 A:Accession: S71713
 A:Molecule type: DNA
 A:Residues: 1-150 <GAW>
 A:Cross-references: EMBL:X92441
 A:Genetics:
 A:Gene: SGD:MGM1
 A:Cross-references: SGD:S0005737; MIPS:YOR211c
 A:Map position: 15R
 A:Genome: nuclear
 C:Function:
 A:Description: mitochondrial genome maintenance
 C:Keywords: GTP binding; mitochondrial; nucleotide binding; P-loop
 F:238-245/Region: nucleotide-binding motif A (P-loop)

Query Match 43.9%; Score 47; DB 2; Length 902;
 Best Local Similarity 72.7%; Pred. No. 19;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 7 KSYFSDCOVST 17
 ||:|||||
 DB 496 KKYFTNCQVST 506

RESULT 5
 S42357
 probable zinc finger protein - common cuttlefish
 C:Species: Sepia officinalis (common cuttlefish)
 C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
 C:Accession: S42357
 R:Martin-Poncheu, A.; Mouters-Tyrou, D.; Pudlo, B.; Buissine, E.; Sautiere, P.
 Eur. J. Biochem. 220, 463-468, 1994
 A:Title: Isolation and characterization of a small putative zinc finger protein from cutt
 A:Reference number: S42357; MUID:94170795; PMID:8125104
 A:Accession: S42357
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-51 <MAR>
 A:Cross-references: UNIPROT:P80304
 C:Keywords: DNA binding; nucleus

C:Superfamily: rice dwarf virus minor outer capsid protein, 127K

Query Match 40.2%; Score 43; DB 2; Length 1148;
 Best Local Similarity 53.8%; Pred. No. 1,le02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 9 YFSDCQVSTFRSV 21
 DB 326 YITECEVDLRSV 338

RESULT 11

S03770
 pectinesterase (EC 3.1.1.11) precursor - *Erwinia chrysanthemi*
 N:Alternate names: pectin methyl esterase
 C:Species: *Erwinia chrysanthemi*

C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 09-Jul-2004
 C:Accession: S03770

R:Plasmid, G.S.

Mol. Microbiol. 2, 247-254, 1988

A:Title: Molecular cloning and nucleotide sequence of the pectin methyl esterase gene of
 A:Reference number: S03770; MUID:88246052; PMID:2837615
 A:Accession: S03770

A:Molecule type: DNA

A:Residues: 1-366 <PLA>

A:Cross-references: UNIPROT:P07863; GB:Y00549; NID:942442; PIDN:CAA66628.1; PID:942443

C:Genetics:

A:Gene: pme

C:Superfamily: pectinesterase

C:Keywords: carboxylic ester hydrolase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-366/Product: pectin esterase #status predicted <MAT>

Query Match 39.3%; Score 42; DB 2; Length 366;
 Best Local Similarity 60.0%; Pred. No. 51;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 7 KSYFSDCQVS 16
 DB 186 RSFSDCRIS 195

RESULT 12

UN0799
 pectinesterase (EC 3.1.1.11) precursor - *Erwinia chrysanthemi*
 N:Alternate names: pectin methyl esterase

C:Species: *Erwinia chrysanthemi*

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004

C:Accession: UN0799; FN0623

R:Laurent, F.; Kotoujansky, A.; Labessee, G.; Bertheau, Y.

Gene 131, 17-25, 1993

A:Title: Characterization and overexpression of the pme gene encoding pectin methyl ester

A:Reference number: UN0799; MUID:93380669; PMID:8370537

A:Accession: UN0799

A:Molecule type: DNA

A:Residues: 1-366 <LAI>

A:Cross-references: UNIPROT:P07863; GB:L07644; NID:9148465; PIDN:AAA24852.1; PID:9148466

A:Experimental source: strain 1937

A:Accession: FN0623

A:Molecule type: protein

A:Residues: 25-34 <LAI>

C:Comment: This enzyme is required for the full development of the soft rot disease of p

C:Genetics:

A:Gene: pem

C:Superfamily: pectinesterase

C:Keywords: carboxylic ester hydrolase

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-366/Product: pectinesterase #status predicted <MAT>

Query Match 39.3%; Score 42; DB 2; Length 366;
 Best Local Similarity 60.0%; Pred. No. 51;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 7 KSYFSDCQVS 16
 DB 186 RSFSDCRIS 195

RESULT 13

E87753
 protein C43B11.5 (imported) - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C:Accession: E87753

R:Anonymous; The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_eleg

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: E87753

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-368 <STO>

A:Cross-references: UNIPROT:P91147; GB:chr_I; PIDN:AAB37621.1; PID:91703567; GSPDB:GN0001

C:Genetics:

A:Gene: C43B11.5

A:Map position: 1

Query Match 39.3%; Score 42; DB 2; Length 368;
 Best Local Similarity 53.8%; Pred. No. 51;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FRNSISKYSFSDC 13
 DB 312 FRNGISITVYQFC 324

RESULT 14

T29779

hypothetical protein C50F2.5 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T29779

R:Du, Z.; Le, T.T.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of *C. elegans* cosmid C50F2.

A:Reference number: Z20684

A:Accession: T29779

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-368 <DUZ>

A:Cross-references: UNIPROT:P91179; EMBL:U80445; PIDN:AAB37799.1; GSPDB:GN00019; CESP:CS

A:Experimental source: strain Bristol N2; clone C50F2

C:Genetics:

A:Gene: CESP:C50F2.5

A:Map position: 1

A:Introns: 27/3; 130/3; 257/3; 342/3

Query Match 39.3%; Score 42; DB 2; Length 368;
 Best Local Similarity 53.8%; Pred. No. 51;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FRNSISKYSFSDC 13
 DB 312 FRNGISITVYQFC 324

RESULT 15

AB0252

glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) [imported] - *Yersinia pestis* (strain CC

C:Species: *Yersinia pestis*

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C:Accession: AB0252

R:Patrick, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.; I

11. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A;Reference number: AB0001; MUID:21470413; PMID:11586360
 A;Accession: AB0252
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-494 <KUR>
 A;Cross-references: UNIPROT:Q8ZET7; GB:AL590842; PIDN:CAC90878.1; PID:G15980077; GSPDB:C
 C;Genetics: zwf
 C;Superfamily: glucose-6-phosphate dehydrogenase
 C;Keywords: oxidoreductase

Query Match 39.3%; Score 42; DB 2; Length 494;
 Best Local Similarity 46.2%; Pred. No. 69;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 NSSIKSYFSPDCQV 15
 | : ||: ||
 Db 163 NDQVAEYFNCCQV 175

Search completed: October 27, 2005, 18:57:26
 Job time : 13.9907 secs

Geoffrey Hamlyn, Esq.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:34:49 ; Search time 61.0465 Seconds
(without alignments)
176.155 Million cell updates/sec

Title: US-10-612-090-16

Perfect score: 107
Sequence: 1 FRNSISIKSYFSDQVSTFRSV 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	1148	Q9H7S7	Q9H7S7 homo sapien
2	107	100.0	6995	Q6R8K2	Q6R8K2 homo sapien
3	107	100.0	22152	Q8WX17	Q8WX17 homo sapien
4	96	89.7	258	Q9D1H1	Q9D1H1 mus musculus
5	54	50.5	496	Q59477	Q59477 pyrococcus
6	51	47.7	368	Q6CS86	Q6CS86 kluyveriomy
7	51	47.7	663	Q7R25	Q7R25 plasmodium
8	51	47.7	955	Q8IDT3	Q8IDT3 plasmodium
9	49	45.8	493	Q9V105	Q9V105 pyrococcus
10	49	45.8	548	Q75103	Q75103 oryza sativ
11	49	45.8	1524	Q7RCM2	Q7RCM2 plasmodium
12	48	44.9	411	Q86128	Q86128 dictyostell
13	48	44.9	809	Q94527	Q94527 schizosacch
14	47	43.9	253	TRPC_BACCR	TRPC_BACCR
15	47	43.9	881	MGM1_YEAST	MGM1_YEAST
16	47	43.9	1235	Q8PF69	Q8PF69 saccharomy
17	46	43.0	253	Q73B09	Q73B09 xanthomonas
18	46	43.0	505	Q8T2W9	Q8T2W9 pyrococcus
19	45	42.1	185	Q8TPM5	Q8TPM5 methanosa
20	45	42.1	511	Q8TPW7	Q8TPW7 methanosa
21	45	42.1	511	Q8TRD8	Q8TRD8 methanosa
22	45	42.1	601	Q9W712	Q9W712 pseudopleur
23	45	42.1	682	Q8EWF4	Q8EWF4 mycoplasma
24	45	42.1	891	Q6FLK2	Q6FLK2 candida gla
25	45	42.1	1812	Q81538	Q81538 plasmodium
26	44.5	41.6	847	Q93XL9	Q93XL9 rosa hybrid
27	44	41.1	51	SPE_SEPOF	SPE_SEPOF
28	44	41.1	131	Q81B89	Q81B89 enterococc
29	44	41.1	145	NUSB_BORBU	NUSB_BORBU
30	44	41.1	174	Q6B365	Q6B365 uncultured
31	44	41.1	187	Q7PMW5	Q7PMW5 anopheles g

32	44	41.1	270	Q757C9	Q757C9 aahbya goss
33	44	41.1	274	Q6LPV1	Q6LPV1 photobacter
34	44	41.1	330	Q8GC33	Q8GC33 leuconostoc
35	44	41.1	583	Q6MMQ0	Q6MMQ0 bdellovibri
36	44	41.1	810	Q8W5J3	Q8W5J3 oryza sativ
37	44	41.1	810	Q7XD67	Q7XD67 oryza sativ
38	44	41.1	933	Q7MG24	Q7MG24 vibrio vuln
39	44	41.1	933	Q8D4J2	Q8D4J2 streptococc
40	44	41.1	1063	Q934I4	Q934I4 streptococc
41	44	41.1	1063	Q6W9D8	Q6W9D8 streptococc
42	43	40.2	122	Q9NA57	Q9NA57 caenorhabdi
43	43	40.2	212	Q8EBH3	Q8EBH3 drosophila
44	43	40.2	253	TRPC_BACAN	TRPC_BACAN
45	43	40.2	253	Q63BC9	Q63BC9 bacillus ce

ALIGNMENTS

RESULT 1
Q9H7S7 PRELIMINARY; PRT; 1148 AA.
ID Q9H7S7
AC Q9H7S7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
OS Hypothetical protein FLJ14303.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta; DOI=10.1038/ng1285;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niinomiya K., Iehibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hirakawa S., Chiba Y., Iehara S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshioka Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamagaki M., Watanabe K., Kumagai A., Itakura S., Fukunumi Y.,
RA Fujimori Y., Komiyama M., Tashiro K., Tanigami A., Fujitawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Moritaga M., Sasaki M.,
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RT Nat. Genet. 36:40-45(2004).
RL EMBL; AK024365; BAB14899.1; -.
DR HSSP; Q9D1H1; IIVZ.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 7.
DR PROSITE; PS50024; SEA; 3.
SQ SEQUENCE 1148 AA; 127957 MW; 3861B0D5EFD8BAC CRC64;

```

Query Match          100.0%; Score 107; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
    |||||
Db 1002 FRNSIKSYFSDCOVSTFRSV 1022

RESULT 2
O96RK2 ID PRELIMINARY; PRT; 6995 AA.
AC O96RK2:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Mucin 16 (Fragment).
GN Name=MUC16;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347965; PubMed=11369781; DOI=10.1074/jbc.M103554200;
RY Yin B.W., Lloyd K.O.;
RT "Molecular cloning of the cal25 ovarian cancer antigen. Identification
as a new mucin, mucin 16."
RT J. Biol. Chem. 276:27371-27375 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX Lloyd K.O., Yin B.W.T.;
RT Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF361486; AK74120.3; -.
DR HSSP; Q9D1H1; 11VZ.
DR InterPro; IPR00194; ATPase_a/bcentre.
DR InterPro; IPR00082; SEA.
DR Pfam; PF01390; SEA; 20.
DR SMART; SM00200; SEA; 10.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00024; SEA; 6.
RT NON TER
SQ SEQUENCE 6995 AA; 744958 MW; 80C797DBDBF33A2B CRC64;

Query Match          100.0%; Score 107; DB 2; Length 6995;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
    |||||
Db 6849 FRNSIKSYFSDCOVSTFRSV 6869

RESULT 3
O8WX17 ID PRELIMINARY; PRT; 22152 AA.
AC O8WX17:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2004 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Ovarian cancer related tumor marker CA125.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21646939; PubMed=11786729; DOI=10.1159/000050638;
RY O'Brien T.J., Beard J.B., Underwood L.J., Dennis R.A., Santin A.D.,
RT "The CA 125 gene: an extracellular superstructure dominated by repeat
sequences."
RT Tumour Biol. 22:348-366 (2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RY O'Brien T.J., Underwood L.J., Beard J.B.;
RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144442; AAL6513.2; -.
DR GeneW; HGNC:1582; MUC16.
DR Pfam; PF01390; SEA; 51.
DR SMART; SM00200; SEA; 23.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_2.
DR PROSITE; PS00024; SEA; 11.
SQ SEQUENCE 22152 AA; 235268 MW; B3EBDF19997A440 CRC64;

Query Match          100.0%; Score 107; DB 2; Length 22152;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 FRNSIKSYFSDCOVSTFRSV 21
    |||||
Db 22006 FRNSIKSYFSDCOVSTFRSV 22026

RESULT 4
O9D1H1 ID PRELIMINARY; PRT; 258 AA.
AC O9D1H1:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched
library, clone:1110008114 product:hypothetical SEA domain containing
protein, full insert sequence.
GN Name=1110008114R1k;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RT Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RT Nature 420:563-573 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes."
RT Genome Res. 10:1617-1630 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuai T., Teshiro H., Itoh M.,

```

```
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsuki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Mutamatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.",
RL Genome Res. 10:1757-1771(2000).
RN [6]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carinici P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imorani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Teijima Y., Taya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC003577; BAB22869.1; -.
DR PDB; 1IV2; NMR; A=60-190.
DR MGD; MGI:1920982; 110008114Rik.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
DR KX Hypothetical protein.
SQ SEQUENCE 258 AA; 29425 MW; B64D9B6394D84E7 CRC64;

Query Match 89.7%; Score 96; DB 2; Length 258;
Best Local Similarity 90.5%; Pred. No. 5.8e-07;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVSTFRSV 21
Db 111 FRNSIKSYFSDCOVLAFRSV 131

RESULT 5
059477 PRELIMINARY; PRT; 496 AA.
AC 059477;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 496aa long hypothetical TRK system potassium uptake protein.
GN Orderedocunames=PH1813;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=OT3;
RA MEDLINE=98344137; PubMed=9679194;
RA Kawababayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takaiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maenuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.",
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAB30932.1; -.
DR PIR; E71192; E71192.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008324; F:cation transporter activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR InterPro; IPR003445; Cat_transpt.
```

```
DR InterPro; IPR004772; K_transpttrk.
DR Pfam; PF02386; TrKH; 1.
DR TIGRPFAM; TIGR00933; 2a38; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 496 AA; 55062 MW; 599AFA1CA8711EA2 CRC64;

Query Match 50.5%; Score 54; DB 2; Length 496;
Best Local Similarity 58.8%; Pred. No. 5.1;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 FRNSIKSYFSDCOVST 17
Db 270 FKNRSIRSFMDIQVT 286

RESULT 6
ID 06CS86 PRELIMINARY; PRT; 368 AA.
AC 06CS86;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Kluyveromyces lactic strain NRRL Y-1140 chromosome D of strain NRRL Y-
DE 1140 of Kluyveromyces lactic.
GN ORFNames=KLIA0D03036g;
OS Kluyveromyces lactic NRRL Y-1140.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=284590;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neugegiste C., Talla E.,
RA Goffard N., Frangul L., Aigle M., Anhouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistrane A., Boyer J., Catolico L., Confanier F., de Darvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Gropi A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Olier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Louvel M., Westhof B., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Pukhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts." ;
RL Nature 430:35-44(2004).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382124; CAH00299.1; -.
DR InterPro; IPR007852; Cdc73.
DR Pfam; PF05179; CDC73; 1.
DR SEQUENCE 368 AA; 42023 MW; 3A6E4A8F6647A108 CRC64;

Query Match 47.7%; Score 51; DB 2; Length 368;
Best Local Similarity 45.0%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 RNSSIKSYFSDCOVSTFRSV 21
Db 72 RNSSIKSYFSDCOVSTFRSV 91

RESULT 7
ID 07RF25 PRELIMINARY; PRT; 663 AA.
AC 07RF25;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
```

DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DB Hypothetical protein.
 GN Name=PY04556;
 OS Plasmodium yoelii yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kocot T.W., Petrea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengut J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 Shalton S.J., van Aken S.E., Riedmiller S.B., Feldblum T.V.,
 Cho J.K., Quekeshubush J., Sedegah M., Shoabdi A., Cummings L.M.,
 Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Vaideya A.B.,
 van Lin L.H., Janse C.U., Waters A.P., Smith H.O., White O.R.,
 Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002). The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; AABL01001396; EAA16437.1; -.
 DR Hypothetical protein.
 KW SEQUENCE 663 AA; 79594 MW; 0A42FCE190F415B0 CRC64;

Query Match 47.7% Score 51; DB 2; Length 663;
 Best Local Similarity 53.3%; Pred. No. 21;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Oy 6 IKSYPSDCQVSTFRS 20
 Db 458 LKRYIECOVPTFRS 472
 ||:|||||:
 ||:|||||:

RESULT 8
 ID 081DT3 PRELIMINARY; PRT; 955 AA.
 AC 081DT3;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical protein PF13_0219.
 GN Name=PF13_0219;
 OS Plasmodium falciparum (Isolate 3D7).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
 Barrington M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844509; CAD52536.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 955 AA; 114349 MW; 10AB720DBCEFC999 CRC64;

Query Match 47.7% Score 51; DB 2; Length 955;
 Best Local Similarity 41.2%; Pred. No. 30;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 Oy 4 SIKSYFSDCQVSTFRS 20
 Db 771 SYFKYVAECQIOTFRN 787
 ||:|||||:
 ||:|||||:

RESULT 9
 Q9V105

ID Q9V105 PRELIMINARY; PRT; 493 AA.
 AC Q9V105;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE TRK/H trk potassium uptake system protein.
 GN ORFNames=PAB0248;
 OS Pyrococcus abyssi.
 CC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OK NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GES / Ozeay;
 RX MEDLINE=22511545; PubMed=12622808;
 RA Cohen G.N., Barbe V., Flamant D., Galperin M., Hellig R., Lecompte O.,
 Poon O., Priour D., Querellou J., Rapp R., Thierry J.-C.,
 Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
 RT "An integrated analysis of the genome of the hyperthermophilic
 archaeon Pyrococcus abyssi.";
 RL Mol. Microbiol. 47:1495-1512(2003).
 DR EMBL; AJ248284; CAB49294.1; -.
 DR PIR; G75151; G75151.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0008324; F:cation transporter activity; IEA.
 DR GO; GO:0008124; P:cation transport; IEA.
 DR InterPro; IPR003445; Cat. transp.
 DR InterPro; IPR004772; K. transp. Trk.
 DR Pfam; PF02386; TrkH; 1.
 DR TIGRFAMs; TIGR00933; 2a38; 1.
 KW Complete proteome.
 SO SEQUENCE 493 AA; 54741 MW; 0B26B7B2B5F81283 CRC64;

Query Match 45.8% Score 49; DB 2; Length 493;
 Best Local Similarity 47.1%; Pred. No. 32;
 Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 FRNSIKSYFSDCQVST 17
 Db 266 FRDKSLRNFPKDIQVKT 282
 ||:|||||:
 ||:|||||:

RESULT 10
 ID 075L03 PRELIMINARY; PRT; 548 AA.
 AC 075L03;
 DT 05-JUN-2004 (TReMBLrel. 27, Created)
 DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
 DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE 'contains similar to RNA recognition motif. (A.k.a. RRM, RBD, or RNP
 DE domain), PF00076' (Putative RNA recognition motif (RRM)-containing
 DE protein).
 GN Name=OSUNB0041A22.21; Synonyms=OJ1654_B10.6;
 OS Oryza sativa (Japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriaristoidaeae; Oryzaeae; Oryza.
 OK NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chow T.-Y., Heing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
 Chao Y.-T., Lee P.-F., Chang S.-J., Chen H.-C., Chen S.-K.,
 Chen T.-R., Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y.,
 Hsiao S.-H., Hsiung J.-N., Hsu C.-H., Kau P.-I., Lee M.-C.,
 Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
 Wu H.-P., Shaw J.-F.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

RESULT 11
 ID 075L03 PRELIMINARY; PRT; 548 AA.
 AC 075L03;
 DT 05-JUN-2004 (TReMBLrel. 27, Created)
 DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
 DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE 'contains similar to RNA recognition motif. (A.k.a. RRM, RBD, or RNP
 DE domain), PF00076' (Putative RNA recognition motif (RRM)-containing
 DE protein).
 GN Name=OSUNB0041A22.21; Synonyms=OJ1654_B10.6;
 OS Oryza sativa (Japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriaristoidaeae; Oryzaeae; Oryza.
 OK NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chow T.-Y., Heing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
 Chao Y.-T., Lee P.-F., Chang S.-J., Chen H.-C., Chen S.-K.,
 Chen T.-R., Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y.,
 Hsiao S.-H., Hsiung J.-N., Hsu C.-H., Kau P.-I., Lee M.-C.,
 Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
 Wu H.-P., Shaw J.-F.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.


```
RA L4 Y.-F., Lin S.-D., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,  
  Wu H.-P., Shaw J.-F.;  
  "Oryza sativa BAC OJ1654 B10 genomic sequence."  
  Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
  DR EMBL: AC093921; AAC75242.1; -;  
  DR EMBL: AC108504; AA044075.1; -;  
  DR GO: GO:0003676; F:nucleic acid binding; IEA.  
  DR InterPro: IPR006650; A/AMP deam AS.  
  DR InterPro: IPR000504; RNA rec mot.  
  DR InterPro: IPR001878; ZnF_CCHC.  
  DR Pfam: PF00076; RRM_1.2.  
  DR Pfam: PF00098; ZF_CCHC; 4.  
  DR PRINTS: PR00939; C2HCZNFINGER.  
  DR SMART: SM00360; RRM; 2.  
  DR SMART: SM00343; ZnF_C2HC; 4.  
  DR PROSITE: PS00485; A_DEAMINASE; UNKNOWN_1.  
  DR PROSITE: PS0102; RRM; 2.  
  DR PROSITE: PS0158; ZF_CCHC; 4.  
  DR PROSITE: PS0158; ZF_CCHC; 4.  
  SQ SEQUENCE 548 AA; 60337 MW; 5B71485E26025D75 CRC64;  
  
  Query Match 45.8%; Score 49; DB 2; Length 548;  
  Best Local Similarity 50.0%; Pred. No. 35;  
  Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
  QY 6 IKSYFSDCQVSTFR 19  
  DB 283 LKKFFSDCKISSIR 296  
  
  RESULT 11  
  QTRCM2 PRELIMINARY; PRT; 1524 AA.  
  ID QTRCM2;  
  AC QTRCM2;  
  DT 01-MAR-2004 (TRENBLrel. 26, Created)  
  DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
  DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
  DE Hypothetical protein (Fragment).  
  GN Name=PY05757;  
  OS Plasmodium Yoelli Yoelli.  
  OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
  NC NCBITaxId=73239;  
  RN (1)  
  RP SEQUENCE FROM N.A.  
  RC STRAIN=17XNL;  
  RX PubMed=1236885; DOI=10.1038/nature01099;  
  RA Carlton J.M., Anguioni S.V., Suh B.B., Kooij T.W., Petrea M.,  
  Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
  Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
  Shalton S.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,  
  Florens J.K., Quackenbush J., Sedegah M., Shoish A., Cummings L.M.,  
  Florens J., Yates F.R., III, Raine J.D., Sinden R.E., Harris M.A.,  
  Cunningham D.A., Preiser P.R., Bergman L.W., Vaideya A.B.,  
  van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,  
  Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
  Carucci D.J.;  
  "Genome sequence and comparative analysis of the model rodent malaria  
  parasite Plasmodium yoelli yoelli."  
  Nature 419:512-519 (2002).  
  -!- CAUTION: The sequence shown here is derived from an  
  EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
  preliminary data.  
  CC EMBL: AABL0100186; EAA17831.1; -;  
  CC EMBL: AABL0100186; EAA17831.1; -;  
  DR Hypothetical protein.  
  KW NON_TER  
  FT  
  SQ SEQUENCE 1524 AA; 178163 MW; E2E6D84B9E8FBAFD CRC64;  
  
  Query Match 45.8%; Score 49; DB 2; Length 1524;  
  Best Local Similarity 44.4%; Pred. No. 1e+02;  
  Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
  
  QY 1 FNNSSIKSYFSDCQVSTFR 18  
  DB 1152 FVNDSTRSTLSECFNSF 1169
```

```
RESULT 12  
ID Q86128 PRELIMINARY; PRT; 411 AA.  
AC Q86128;  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Similar to Dictyostelium discoidium (slime mold). Mpa protein.  
OS Dictyostelium discoidium (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
NC NCBITaxId=44689;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RX MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;  
RA Gloeckner G., Eichinger L., Szafianski K., Pachbat J., Dear P.,  
  Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
  Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
  "Sequence and analysis of chromosome 2 of Dictyostelium discoidium."  
  Nature 418:79-85 (2002).  
  RT  
  RL Nature 418:79-85 (2002).  
  RN (2)  
  RP SEQUENCE FROM N.A.  
  RC STRAIN=AX4;  
  RA Baumgart C.;  
  RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
  DR EMBL: AC116305; AA052322.1; -;  
  SQ SEQUENCE 411 AA; 46427 MW; A0974D03F43FB6F1 CRC64;  
  
  Query Match 44.9%; Score 48; DB 2; Length 411;  
  Best Local Similarity 38.1%; Pred. No. 37;  
  Matches 8; Conservative 7; Mismatches 4; Indels 2; Gaps 1;
```

```
QY 1 FNNSSIKSYFSDCQVSTFRSV 21  
DB 229 FNNSSIKSYFSDCQVSTFRSV 247  
  
RESULT 13  
ID Q94527 PRELIMINARY; PRT; 809 AA.  
AC Q94527;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE SPBC609.03 protein.  
GN Name=SPBC609.03;  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
NC NCBITaxId=4896;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=9728-;  
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature9724;  
  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
  Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,  
  Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
  Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
  James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,  
  Mooney P., Moule S., Pearson D., Quail M.A., Rabinowitch E.,  
  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
  Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,  
  Woodward J., Wolke G., Aert R., Robben J., Grynolprez B.,  
  Weljens I., Vanstreels E., Rieger M., Schater M., Muller-Auer S.,  
  Gabel C., Fuchs M., Dusterhoft A., Fritz C., Holzer E., Moestl D.,  
  Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
```

```

RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wandutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaie V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe."
RA Nature 415:871-880(2002).
RL
CC -1- SIMILARITY: Contains 5 MD repeats.
CC EMBL: AL035226; CAA22832.1; -.
CC PIR: T40574; T40574.
CC GeneDB: SPombe; SPBC609.03; -.
CC InterPro: IPR001680; WD40.
CC InterPro: IPR011046; WD40_like.
CC Pfam: PF00400; WD40; 5.
CC PRINTS: PR00320; GPROTEINRPT.
CC PRODOM: PD000018; WD40; 1.
CC SMART: SM00320; WD40; 6.
CC PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
CC PROSITE: PS00082; WD_REPEATS_2; 1.
CC PROSITE: PS50294; WD_REPEATS_REGION; 2.
CC Repeat: WD repeat.
SQ SEQUENCE 809 AA; 92498 MW; 4877443546BD057E CRC64;

Query Match 44.9%; Score 48; DB 2; Length 809;
Best Local Similarity 36.8%; Pred. No. 76;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 3 NSSISYSPDCOVSTFRSV 21
DB 658 NTHVKSYYGHCVESIKNV 676

RESULT 14
TRPC_BACCR STANDARD; PRT; 253 AA.
AC
Q81GQ7;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Inole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS).
DN Name=trpc; OrderedLocustNames=BC1235;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=226900;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapural V., Bhattacharya A., Resnik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goldstein E., Larsen N., D'Souza M., Malunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fontein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.C.;
RA "Genome sequence of Bacillus cereus and comparative analysis with
RA Bacillus anthracis."
RA Nature 423:87-91(2003).
RL
CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-
CC phosphate + C(11)-(3-indolyl)-glycerol 3-phosphate + CO(2) + H(2)O.
CC -1- PATHWAY: tryptophan biosynthesis; fourth step.
CC -1- SIMILARITY: Belongs to the trpc family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE017001; AAP08220.1; -.
CC HSPF: P00909; IPT1.

```

```

DR HAMAP; MF_00134; -, 1.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR001468; IGPS.
DR InterPro; IPR011060; Ribp_bind_barrel.
DR Pfam; PF00218; IGPS; 1.
DR PRODOM; PD00511; IGPS; 1.
DR PROSITE; PS00614; IGPS; 1.
DR Complete proteome; Decarboxylase, lyase, tryptophan biosynthesis.
SQ SEQUENCE 253 AA; 28289 MW; 2519649E1A8081C8 CRC64;

Query Match 43.9%; Score 47; DB 1; Length 253;
Best Local Similarity 57.1%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 NSSISYSPDCOV 16
DB 239 SSSISFFEDCKVN 252

RESULT 15
MGM1 YEAST STANDARD; PRT; 881 AA.
AC P32266; Q02609; Q08627;
DT 01-OCT-1993 (Rel. 27, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE MGM1 protein, mitochondrial precursor.
DN Name=MGM1; OrderedLocustNames=YOR211C; ORFNames=YOR50-1;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bj1-8C;
RC MEDLINE=92192451; PubMed=1532158;
RA Jones B.A., Fangman W.L.;
RA "Mitochondrial DNA maintenance in yeast requires a protein containing
RA a region related to the GTP-binding domain of dynamin."
RA Genes Dev. 6:380-389(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93365024; PubMed=7916673;
RA Guan K., Fair L., Marshall T., Deschenes R.J.;
RA "Nuclear mitochondrial structure and genome maintenance in yeast
RA requires the dynamin-like product of the MGM1 gene."
RA Curr. Genet. 24:141-146(1993).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313270; PubMed=9169874;
RA Dujon B., Albermann K., Alder M., Alexandraki D., Ansgore W.,
RA Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonie R.,
RA Boyer J., Camasses A., Casamayor A., Casas C., Cheret G.,
RA Czepluch C., Daigman-Fornier B., Dang D.V., de Haan M., Delius H.,
RA Durand P., Fairhead C.A., Feldman H., Gallon L., Galisson F.,
RA Gamo F.-J., Gancedo C., Goffeau A., Goulding S.E., Grivell L.A.,
RA Herando B., Hand N.J., Hani J., Hattenhorst U., Hebling U.,
RA Hernandez Y., Herrero E., Heumann K., Hiesel R., Hilger F., Hofmann B.,
RA Hollenberg C.P., Hughes B., Janniaux J.-C., Kalogeropoulos A.,
RA Katsoulou C., Kordes E., Lafuente M.J., Landt O., Louis E.J.,
RA Maatse A.C., Madania A., Manuault G., Marck C., Martin R.P.,
RA Mewes H.-W., Michaux G., Paces V., Parle-McCormick A.G., Pearson B.M.,
RA Perrin A., Petersson B., Poch O., Pohl T.M., Polley R.,
RA Portetelle D., Pujol A., Purnelle B., Ramezani Rad M., Rechmann S.,
RA Schwager C., Schweizer M., Sor F., Sterly F., Tarasov I.A.,
RA Teodoru C., Tettelin H., Thierry A., Tobiasch E., Tzermia M.,
RA Uhlen M., Unselid M., Valens M., Vandenbol M., Vetter I., Vleck C.,
RA Voet M., Volckaert G., Voss H., Wandutt R., Wedler H., Wiemann S.,
RA Winsor B., Wolfe K.H., Zollner A., Zumestein E., Klein K.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XV."
RA Nature 387:98-102(1997).
[4]
RP SEQUENCE OF 1-784 FROM N.A.

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62834; CA444637.1; ALT FRAME.
DR EMBL; L07419; -. NOT ANNOTATED CDS.
DR EMBL; Z75119; CAA99426.1; ALT_INIT.
DR EMBL; Z75120; CAA99428.1; ALT_INIT.
DR EMBL; X92441; CAA63174.1; ALT_INIT.
DR EMBL; L11274; AAB59316.1; -.
DR PIR; S33918; S33918.
DR GenBank; 143799; -.
DR SGD; S000005737; MGM1.
DR GO; GO:0005758; C:mitochondrial intermembrane space; IDA.
DR GO; GO:0008053; P:mitochondrial fusion; IDA.
DR GO; GO:0000002; P:mitochondrial genome maintenance; IDA.
DR GO; GO:0007006; P:mitochondrial membrane organization and bio. . .; IDA.
DR GO; GO:0000001; P:mitochondrion inheritance; IDA.
DR InterPro; IPR01401; DYNamin.
DR Pfam; PF00350; DYNamin_N/1.
DR PRINTS; PRO0195; DYNAMIN.
DR SMART; SM00053; DYNC; 1.
DR PROSITE; PS00410; DYNAMIN; 1.
KW Direct protein sequencing; GTP-binding; Hydrolase; Mitochondrion;
KW Motor protein; Transil peptide.
FT TRANSIT 1 60 Mitochondrion.
FT CHAIN 61 881 MGM1 protein, isoform 1.
FT DOMAIN 149 178 Asp-rich (acidic).
FT NP_BIND 217 224 GTP (Potential).
FT NP_BIND 317 321 GTP (Potential).
FT NP_BIND 385 388 GTP (Potential).
FT VARIANT 294 294 E -> K (in mem17).
FT MUTAGEN 1 1 M->A: Abolishes translation.
FT MUTAGEN 39 39 M->A: No effect on translation.
FT MUTAGEN 81 81 M->A: No effect on translation.
FT MUTAGEN 92 92 M->A: No effect on translation.
FT MUTAGEN 223 223 K->A: Loss of stability.
FT MUTAGEN 224 224 S->N: Loss of GTPase activity.
FT MUTAGEN 224 224 T->A: Loss of GTPase activity.
FT MUTAGEN 824 824 R->A: Loss of GBD function.
FT MUTAGEN 854 854 K->A: Loss of GBD function.
FT CONFLICT 4 4 S -> T (in Ref. 1).
FT CONFLICT 129 129 G -> C (in Ref. 1).
FT CONFLICT 170 170 E -> A (in Ref. 2).
SQ SEQUENCE 881 AA; 99177 MW; 31524B40C5350096 CRC64;
Query Match 43.9%; Score 47; DB 1; Length 881;
Best Local Similarity 72.7%; Prd. No. 1,2e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 7 KSYFSDQVST 17
| : : : :
Db 475 KKRYFNQVST 485

```

This Page Blank (100%)

This Page Blank (100%)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:31:33 : Search time 79.5907 Seconds
(without alignments)
111.765 Million cell updates/sec

Title: US-10-612-090-17
Perfect score: 125
Sequence: 1 SVPNRHTGVDSLINFSPILARRV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	100.0	23	8	ADJ66744 Human Muc
2	125	100.0	108	8	ADJ66729 Human Muc
3	125	100.0	108	8	ADJ66740 Human Muc
4	125	100.0	174	7	ADM30756 Human can
5	125	100.0	178	5	ABP30979 Predicted
6	125	100.0	178	5	ADA08642 Human O77
7	125	100.0	178	7	ADF08985 Secreted
8	125	100.0	233	5	ABP30978 Extracell
9	125	100.0	233	7	ADA08641 Human O77
10	125	100.0	233	7	ADF08984 Secreted
11	125	100.0	284	6	ABU54859 Human CA1
12	125	100.0	318	5	ABP31025 Antno aci
13	125	100.0	318	7	ADA08747 Human O77
14	125	100.0	318	7	ADF09090 Secreted
15	125	100.0	367	4	AAE12631 Human gen
16	125	100.0	438	3	AAE12555 Human ova
17	125	100.0	438	4	AAE12623 Human gen
18	125	100.0	438	5	ABP30973 Partial p
19	125	100.0	438	5	ABP30899 0772P c10
20	125	100.0	438	7	ADA08636 Human ova
21	125	100.0	438	7	ADA08543 Human ova
22	125	100.0	438	7	ADF08886 Secreted
23	125	100.0	438	7	ADF08979 Secreted
24	125	100.0	438	7	ADG46173 Human ova
25	125	100.0	439	6	ABU54709 Human CA1

26	125	100.0	526	4	AAE12634 Human gen
27	125	100.0	748	8	ADS94302 CA 125/O7
28	125	100.0	809	8	ADS94303 CA 125/O7
29	125	100.0	833	3	AAE12554 Human ova
30	125	100.0	833	5	ABP30898 Human ova
31	125	100.0	833	7	ADA08542 Human ova
32	125	100.0	833	7	ADF08885 Secreted
33	125	100.0	833	7	ADG46172 Human ova
34	125	100.0	914	3	AAE12552 Human ova
35	125	100.0	914	4	AAE12553 Human ova
36	125	100.0	914	5	ABP30968 Hypochet1
37	125	100.0	914	5	ABP30896 0772P p10
38	125	100.0	914	7	ADA08631 Human O77
39	125	100.0	914	7	ADA08465 Human ova
40	125	100.0	914	7	ADF08974 Secreted
41	125	100.0	914	7	ADF08808 Secreted
42	125	100.0	914	7	ADG46095 Human ova
43	125	100.0	914	8	ADN40451 Human bre
44	125	100.0	1148	4	AAE12552 Human ova
45	125	100.0	1148	4	ABE50283 HOST-1 ov

ALIGNMENTS

RESULT 1

ADJ66744 standard; peptide; 23 AA.

ADJ66744; (first entry)
06-MAY-2004
Human Muc1/Muc16 protein-related peptide SegID17.
monoclonal antibody; epitope; non-shed extracellular portion;
shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
breast cancer; ovarian cancer.
Homo sapiens.
WO2004005470-A2.
15-JAN-2004.
03-JUL-2003; 2003WO-US020907.
PR 03-JUL-2002; 2002US-0393094P.
(IMMU-) IMMUNOGEN INC.
Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
Water CA;
WPI; 2004-091350/09.
New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
treating or monitoring malignancies, such as breast or ovarian cancer.
Claim 14; SEQ ID NO 17; 113pp; English.
This invention relates to a novel isolated monoclonal antibody that
specifically binds to an epitope of a non-shed extracellular portion of a
shed antigen or of human Muc1 or Muc16 protein. The invention may be
useful for the development of compounds with a cytostatic or for gene
therapy. The composition and methods are useful in diagnosing, treating
or monitoring malignancies, such as breast or ovarian cancer. The present
sequence is that of a peptide which was used for raising antibodies to
the extracellular, non-shed region of Muc16 and which is claimed in the
specification.
Sequence 23 AA;

```

Query Match          100.0%; Score 125; DB 8; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.9e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1  SVPNRHHTGVDSLGNFSPPLARRV 23
Db      1  SVPNRHHTGVDSLGNFSPPLARRV 23

RESULT 2
ADJ66729
ID      ADJ66729 standard; protein; 108 AA.
XX
AC      ADJ66729;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Human Muc16 epitope amino acid sequence.
XX
KW      monoclonal antibody; epitope; non-shed extracellular portion;
KW      shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
KW      breast cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2004005470-A2.
XX
PD      15-JAN-2004.
XX
PF      03-JUL-2003; 2003WO-US020907.
XX
PR      03-JUL-2002; 2002US-0393094P.
XX
PA      (IMMU-) IMMUNOGEN INC.
XX
PI      Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
PI      Vater CA;
XX
DR      WPI; 2004-091350/09.
XX
NR      New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
PT      treating or monitoring malignancies, such as breast or ovarian cancer.
XX
PS      Claim 13; SEQ ID NO 2; 113pp; English.
XX
CC      This invention relates to a novel isolated monoclonal antibody that
CC      specifically binds to an epitope of a non-shed extracellular portion of a
CC      shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC      useful for the development of compounds with a cytostatic or for gene
CC      therapy. The composition and methods are useful in diagnosing, treating
CC      or monitoring malignancies, such as breast or ovarian cancer. The present
CC      sequence is that of a Muc epitope of the invention.
XX
SQ      Sequence 108 AA;

Query Match          100.0%; Score 125; DB 8; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1  SVPNRHHTGVDSLGNFSPPLARRV 23
Db      37  SVPNRHHTGVDSLGNFSPPLARRV 59

RESULT 3
ADJ66740
ID      ADJ66740 standard; protein; 108 AA.
XX
AC      ADJ66740;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Human Muc16 GST fusion protein amino acid sequence SeqID13.

```

```

XX      monoclona antibody; epitope; non-shed extracellular portion;
XX      shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
XX      breast cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2004005470-A2.
XX
PD      15-JAN-2004.
XX
PF      03-JUL-2003; 2003WO-US020907.
XX
PR      03-JUL-2002; 2002US-0393094P.
XX
PA      (IMMU-) IMMUNOGEN INC.
XX
PI      Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K;
PI      Vater CA;
XX
DR      WPI; 2004-091350/09.
XX
NR      New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
PT      treating or monitoring malignancies, such as breast or ovarian cancer.
XX
PS      Disclosure; SEQ ID NO 13; 113pp; English.
XX
CC      This invention relates to a novel isolated monoclonal antibody that
CC      specifically binds to an epitope of a non-shed extracellular portion of a
CC      shed antigen or of human Muc1 or Muc16 protein. The invention may be
CC      useful for the development of compounds with a cytostatic or for gene
CC      therapy. The composition and methods are useful in diagnosing, treating
CC      or monitoring malignancies, such as breast or ovarian cancer. The present
CC      sequence is that of the human Muc16 GST fusion protein which was used in
CC      the exemplification of the invention.
XX
SQ      Sequence 108 AA;

Query Match          100.0%; Score 125; DB 8; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1  SVPNRHHTGVDSLGNFSPPLARRV 23
Db      37  SVPNRHHTGVDSLGNFSPPLARRV 59

RESULT 4
ADM30756
ID      ADM30756 standard; protein; 174 AA.
XX
AC      ADM30756;
XX
DT      20-MAY-2004 (first entry)
XX
DE      Human cancer linked protein referred to as 192292 SeqID 4.
XX
KW      human; cancer-related gene; screening assay; immuno-conjugate;
KW      cytostatic; immunostimulant; cancer; ovarian cancer.
XX
OS      Homo sapiens.
XX
PN      WO2003075854-A2.
XX
PD      18-SEP-2003.
XX
PF      07-MAR-2003; 2003WO-US007147.
XX
PR      07-MAR-2002; 2002US-0362527P.
XX
PA      (AVAL-) AVALON PHARM.
XX
PI      Young PE, Edner R, Strovel JW;

```

XX WPI; 2003-748320/70.
DR N-PSDB; ADM30755.
XX
PT Identifying an agent that modulates the activity of any of three specific
PT cancer-related genes, potentially useful in treating (ovarian) cancer,
PT comprises detecting a difference in expression of the gene in the
PT presence of the agent.
XX
PS Claim 11; SEQ ID NO 4; 57bp; English.
XX
CC This invention relates to a novel method of identifying an agent that
CC modulates the activity of a cancer-related gene. Specifically, it refers
CC to a screening assay to identify potential antitumor agents, as well as
CC methods to assess the cancerous state of a cell. The present invention
CC describes antibodies against the expression products of these cancer-
CC related genes that are capable of targeting a cancerous cell in vivo.
CC Furthermore it provides immuno-conjugates containing such antibodies,
CC which can be used to deliver target therapeutics thereto. Accordingly,
CC these compositions that exhibit cytostatic and immunostimulant activities
CC are useful for preventing or treating cancer in humans and in particular
CC the treatment of ovarian cancer. This polypeptide sequence is a protein
CC encoded by a human cancer-linked gene of the invention.
XX
SQ Sequence 174 AA;
XX
Query Match 100.0%; Score 125; DB 7; Length 174;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SVPNRHHTGVDSLGNFSPPLARRV 23
Db 47 SVPNRHHTGVDSLGNFSPPLARRV 69
XX
RESULT 5
ABP30979
ID ABP30979 standard; protein; 178 AA.
XX
AC ABP30979;
XX
DT 02-JUL-2002 (first entry)
XX
DE Predicted extracellular domain of 0772P.
XX
KM Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
OS Homo sapiens.
XX
PN WO200206317-A2.
XX
PD 24-JAN-2002.
XX
PF 17-JUL-2001; 2001WO-US022635.
XX
PI 17-JUL-2000; 2000US-00617747.
XX
PR 10-ANG-2000; 2000US-00635801.
XX
PR 20-SEP-2000; 2000US-00667857.
XX
PR 04-APR-2001; 2001US-00827271.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
DR WPI; 2002-164781/21.
XX
PT Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
PS Example 13; Page 361-362; 408bp; English.

XX This invention relates to polypeptides comprising an immunogenic portion
CC of an ovarian carcinoma protein which acts as an immunostimulant and is
CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations and antigen presenting cells that express
CC the polypeptides are useful for stimulating an immune response in a
CC patient and treating ovarian cancer. This sequence represents protein
CC related to the invention
XX
SQ Sequence 178 AA;
XX
Query Match 100.0%; Score 125; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.1e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 SVPNRHHTGVDSLGNFSPPLARRV 23
Db 106 SVPNRHHTGVDSLGNFSPPLARRV 128
XX
RESULT 6
ADA08642
ID ADA08642 standard; protein; 178 AA.
XX
AC ADA08642;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human 0772P partial protein #12.
XX
KM human; gene therapy; ovarian cancer; cancer.
XX
OS Homo sapiens.
XX
PN US2003091580-A1.
XX
PD 15-MAY-2003.
XX
PF 17-JUL-2001; 2001US-00907969.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (MITC/) MITCHAM J L.
PA (KING/) KING G E.
PA (ALGA/) ALGATE P A.
PA (FLIN/) FLING S P.
PA (RETT/) RETTER M W.
PA (FANG/) FANGER G R.
PA (REED/) REED S G.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (HILL/) HILL P.
PA (ALBO/) ALBONE E.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
DR WPI; 2003-532352/50.
XX
PT New isolated 0772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
PT cancer.
XX
PS Disclosure; Page 28; 371bp; English.
XX
CC The invention relates to an isolated 0772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma cDNAs and protein cDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma

CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 125; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 3.1e-13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVPNRHHTGVDSLGNFSPFLARRV 23
 DB 106 SVPNRHHTGVDSLGNFSPFLARRV 128
 RESULT 7
 ID ADF08985 standard; protein; 178 AA.
 XX ADF08985;
 AC
 XX 12-FEB-2004 (first entry)
 DT
 XX Secreted ovarian carcinoma antigen seqid 489.
 DE
 XX gene therapy; protein therapy; vaccine; antibody inhibition;
 KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
 XX secreted ovarian carcinoma antigen.
 XX
 OS Homo sapiens.
 XX
 PN US2003124140-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 17-JUL-2002; 2002US-00198053.
 XX
 PR 17-DEC-1998; 98US-002155681.
 XX 17-DEC-1998; 98US-00216003.
 PR 23-JUN-1999; 99US-00338933.
 PR 24-SEP-1999; 99US-00404879.
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00864441.
 PR 17-JUL-2001; 2001US-00907959.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Bangur CS, Retter MW, Fanger GR, Hill P;
 DR WPI; 2003-897152/82.
 XX
 PT Oncogenic nucleic acids useful for the prevention, diagnosis and
 XX treatment of breast cancer.
 PT
 XX Example 13; SEQ ID NO 489; 399pp; English.
 XX
 CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancer. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patients own production of (II).
 CC Additionally, (II) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigen in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and
 CC antagonists may be used to regulate expression and activity and as

CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunosassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 125; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 3.1e-13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVPNRHHTGVDSLGNFSPFLARRV 23
 DB 106 SVPNRHHTGVDSLGNFSPFLARRV 128
 RESULT 8
 ID ABP30978 standard; protein; 233 AA.
 XX ABP30978;
 AC
 XX 02-JUL-2002 (first entry)
 DT
 XX Extracellular and transmembranal regions of 0772P.
 DE
 XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
 KW
 XX Homo sapiens.
 OS
 XX WO200206317-A2.
 PN
 PD 24-JAN-2002.
 XX
 PF 17-JUL-2001; 2001WO-US022635.
 XX
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
 PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
 XX WPI; 2002-164781/21.
 DR
 XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
 PT protein or its variants, useful for stimulating an immune response in a
 PT patient and treating ovarian cancer.
 PT
 XX Example 13; Page 361; 408pp; English.
 XX
 CC This invention relates to polypeptides comprising an immunogenic portion
 CC of an ovarian carcinoma protein which acts as an immunostimulant and is
 CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations and antigen presenting cells that express
 CC the polypeptides are useful for stimulating an immune response in a
 CC patient and treating ovarian cancer. This sequence represents protein
 CC related to the invention
 CC
 SQ Sequence 233 AA;
 Query Match 100.0%; Score 125; DB 5; Length 233;
 Best Local Similarity 100.0%; Pred. No. 4.2e-13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVPNRHHTGVDSLGNFSPFLARRV 23
 DB 106 SVPNRHHTGVDSLGNFSPFLARRV 128


```

RESULT 9
ADA08641
ID ADA08641 standard; protein; 233 AA.
XX
AC ADA08641;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human O772P partial protein #11.
XX
KM human; gene therapy; ovarian cancer; cancer.
XX
OS Homo sapiens.
XX
PN US2003091580-A1.
XX
PD 15-MAY-2003.
XX
PF 17-JUL-2001; 2001US-00907969.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (MITC/) MITCHAM J L.
PA (KING/) KING G E.
PA (ALGA/) ALGATE P A.
PA (FLIN/) FLING S P.
PA (RETT/) RETTER M W.
PA (FANG/) FANGER G R.
PA (REED/) REED S G.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (HILL/) HILL P.
PA (ALBO/) ALBONE E.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
DR WPI; 2003-532352/50.
XX
PT New isolated O772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
PT cancer.
XX
PS Example 13; Page 28; 371pp; English.
XX
CC The invention relates to an isolated O772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma cDNAs and protein cDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma
CC antigen.
XX
SQ Sequence 233 AA;
XX
Query Match 100.0%; Score 125; DB 7; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVPNRHHGTGVDSLCNFSPPLARRV 23
Db 106 SVPNRHHGTGVDSLCNFSPPLARRV 128
XX
RESULT 10
ADP08984
ID ADP08984 standard; peptide; 233 AA.
XX
AC ADP08984;
XX
DT 12-FEB-2004 (first entry)

```

```

XX
DE Secreted ovarian carcinoma antigen seqid 488.
XX
KM gene therapy; protein therapy; vaccine; antibody inhibition;
XX breast cancer; restorative therapy; diagnostic agent; immunoassay;
XX secreted ovarian carcinoma antigen.
XX
OS Homo sapiens.
XX
PN US2003124140-A1.
XX
PD 03-JUL-2003.
XX
PF 17-JUL-2002; 2002US-00198053.
XX
PR 17-DEC-1998; 98US-00215681.
PR 17-DEC-1998; 98US-00216003.
PR 23-JUN-1999; 99US-00338933.
PR 24-SEP-1999; 99US-00404879.
PR 17-JUL-2000; 2000US-00617747.
PR 10-AUG-2000; 2000US-00636801.
PR 20-SEP-2000; 2000US-00667857.
PR 04-APR-2001; 2001US-00827271.
PR 18-JUN-2001; 2001US-00884441.
PR 17-JUL-2001; 2001US-00907969.
XX
PA (CORI-) CORIXA CORP.
XX
PI Bangur CS, Retter MW, Fanger GR, Hill P;
XX
DR WPI; 2003-897152/82.
XX
PT Oncogenic nucleic acids useful for the prevention, diagnosis and
PT treatment of breast cancer.
XX
PS Example 13; SEQ ID NO 488; 399pp; English.
XX
CC The invention describes nucleic acids (I) and the polypeptides (II) they
CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
CC treating diseases related to their aberrant expression i.e. breast
CC cancers. For example, (I) and (II) may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of (II) by expressing
CC inactive proteins or to supplement the patient's own production of (II).
CC Additionally, (I) may be used to produce (II), by inserting (I) into a
CC host cell and culturing the cell to express the protein (II). (I) And its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. The host cell may also be used as antigens in the production of
CC antibodies against (II) and in assays to identify modulators of (II)'s
CC expression and activity. The anti-(II) antibodies, agonists and
CC antagonists may be used to regulate expression and activity and as
CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
CC immunoassay). This sequence represents a secreted ovarian carcinoma
CC antigen.
XX
SQ Sequence 233 AA;
XX
Query Match 100.0%; Score 125; DB 7; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.2e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 SVPNRHHGTGVDSLCNFSPPLARRV 23
Db 106 SVPNRHHGTGVDSLCNFSPPLARRV 128
XX
RESULT 11
ABU54859
ID ABU54859 standard; protein; 284 AA.
XX
AC ABU54859;

```

```

XX 12-MAR-2003 (first entry)
XX
XX Human CA125 carboxy terminal domain.
DE
XX Human, CA125; protein repeat; chromosome 19q 13.2; amino terminal domain;
XX amino terminal extension; carboxy terminal domain; vaccine; cancer;
XX ovarian cancer; carcinoma.
XX
XX Homo sapiens.
XX
XX WO200283866-A2.
XX
XX 24-OCT-2002.
XX
XX 12-APR-2002; 2002WO-US011734.
XX
XX 17-APR-2001; 2001US-0284175P.
XX 19-JUN-2001; 2001US-0299380P.
XX 27-SEP-2001; 2001US-00965738.
XX 21-DEC-2001; 2001US-0345180P.
XX
XX (UVAR-) UNIV ARKANSAS.
XX
XX O'Brien T, Beard J, Underwood L;
XX
XX WPI; 2003-093013/08.
XX
XX New CA125 molecules, useful as a gold standard for detecting and
XX monitoring the presence of CA125 antigen which can be used for
XX diagnosing, monitoring or treating patients with cancer or for developing
XX vaccine against cancer.
XX
XX Claim 1; Fig 9b; 694pp; English.
XX
XX The invention relates to a CA125 protein comprising: (a) an extracellular
XX amino terminal domain; (b) an amino terminal extension; (c) a multiple
XX repeat domain; and (d) a carboxy terminal domain. The extracellular amino
XX terminal domain comprises 5 genomic exons, the amino terminal extension
XX comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and
XX the carboxy terminal domain comprises a transmembrane anchor with a short
XX cytoplasmic domain, and further comprises 9 genomic exons. The gene for
XX CA125 is located on human chromosome 19q 13.2. Also included are isolated
XX CA125 repeat domains, nucleic acids (including variants, homologues and
XX degenerate versions) encoding CA125 proteins or repeat units, a vector
XX comprising the nucleic acid, a cultured cell comprising the vector, a
XX method of expressing CA125 antigen in a cell, the amino acid sequences of
XX the CA125 repeat units (or their variants, fragments or sequences 50%
XX identical to them), a purified antibody that selectively binds to an
XX epitope in the receptor-binding domain of CA125 antigen (comprising
XX detecting and monitoring the presence of CA125 antigen (comprising
XX recombinant CA125 having at least one repeat unit of the CA125 repeat
XX domain including epitope binding sites), a therapeutic vaccine to treat
XX mammals with elevated CA125 antigen levels or at risk of developing a
XX disease or disease recurrence associated with elevated CA125 antigen
XX levels (comprising recombinant CA125 repeat domains including epitope
XX binding sites) and an antisense oligonucleotide that inhibits the
XX expression of CA125. The CA125 molecule, particularly the multiple repeat
XX domain are useful as a gold standard for detecting and monitoring the
XX presence of CA125 antigen, which can be used for diagnosing, monitoring
XX or treating patients with ovarian cancer and other carcinomas where CA125
XX is expressed. The molecules are also useful for developing a vaccine
XX against cancer. The present sequence is a CA125 repeat protein
XX
SQ Sequence 284 AA;

```

```

RESULT 12
ABP31025
ID ABP31025 standard; protein; 318 AA.
XX
XX AC ABP31025;
XX
XX 02-JUL-2002 (first entry)
XX
XX Amino acid sequence of the 3' constant region of 0772P.
XX
XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
XX Homo sapiens.
XX
XX WO200206317-A2.
XX
XX 24-JAN-2002.
XX
XX 17-JUL-2001; 2001WO-US022635.
XX
XX 17-JUL-2000; 2000US-00617747.
XX 10-AUG-2000; 2000US-00636801.
XX 20-SEP-2000; 2000US-00667857.
XX 04-APR-2001; 2001US-00827271.
XX 18-JUN-2001; 2001US-00884441.
XX
XX (CORI-) CORIXA CORP.
XX
XX Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
XX Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2002-164781/21.
XX
XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
XX protein or its variants, useful for stimulating an immune response in a
XX patient and treating ovarian cancer.
XX
XX Claim 3; Page 399-400; 408pp; English.
XX
XX This invention relates to polypeptides comprising an immunogenic portion
XX of an ovarian carcinoma protein which acts as an immunostimulant and is
XX cytostatic. The polypeptides, polynucleotides, antibodies, fusion
XX proteins, T cell populations and antigen presenting cells that express
XX the polypeptides are useful for stimulating an immune response in a
XX patient and treating ovarian cancer. This sequence represents protein
XX related to the invention
XX
SQ Sequence 318 AA;

```

```

Query Match 100.0%; Score 125; DB 5; Length 318;
Best Local Similarity 100.0%; Pred. No. 66-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHGTGVDLGNFSPPLARRV 23
   |||||
DB 191 SVPNRHHGTGVDLGNFSPPLARRV 213
   |||||

RESULT 13
ADA08747
ID ADA08747 standard; protein; 318 AA.
XX
XX ADA08747;
XX
XX 06-NOV-2003 (first entry)
XX
XX DT Human O772P constant region consensus sequence #1.
XX
XX DE Human O772P constant region consensus sequence #1.
XX
XX KW human; gene therapy; ovarian cancer; cancer.
XX
XX OS Homo sapiens.

```

```

PF      17-JUL-2002; 2002US-00198053.
XX
XX      17-DEC-1998;   98US-00215681.
PR      17-DEC-1998;   98US-00216003.
PR      23-JUN-1999;   99US-00338933.
PR      24-SEP-1999;   99US-00404879.
PR      17-JUL-2000; 2000US-00617747.
PR      10-AUG-2000; 2000US-00636801.
PR      20-SEP-2000; 2000US-00667857.
PR      04-APR-2001; 2001US-00827271.
PR      18-JUN-2001; 2001US-00884441.
PR      17-JUL-2001; 2001US-00907969.
XX
XX
PA      (CORI-) CORIXA CORP.
XX
XX      Bangor CS, Retter MW, Fanger GR, Hill P;
PI
DR      WPI; 2003-897152/82.
DR      N-PSDB; ADF09064.
XX
XX
PT      Oncogenic nucleic acids useful for the prevention, diagnosis and
        treatment of breast cancer.
XX
XX
PS      Example 16; SEQ ID NO 594; 399pp; English.
XX
XX
CC      The invention describes nucleic acids (I) and the polypeptides (II) they
CC      encode. The nucleic acids (I) may be used for preventing, diagnosing and
CC      treating diseases related to their aberrant expression i.e. breast
CC      cancers. For example, (I) and (II) may be used to treat disorders
CC      associated with decreased expression by rectifying mutations or deletions
CC      in a patient's genome that affect the activity of (II) by expressing
CC      inactive proteins or to supplement the patients own production of (II).
CC      Additionally, (I) may be used to produce (II), by inserting (I) into a
CC      host cell and culturing the cell to express the protein (II). (I) and its
CC      complementary sequences may also be used as DNA probes in diagnostic
CC      assays to detect and quantitate the presence of similar nucleic acids in
CC      samples, and therefore which patients may be in need of restorative
CC      therapy. The host cell may also be used as antigens in the production of
CC      antibodies against (II) and in assays to identify modulators of (II)'s
CC      expression and activity. The anti-(II) antibodies, agonists and
CC      antagonists may be used to regulate expression and activity and as
CC      diagnostic agents for detecting the presence of (II) in samples (e.g. by
CC      immunassay). This sequence represents a secreted ovarian carcinoma
CC      antigen.
XX
XX
SQ      Sequence 318 AA;
        Query Match          100.0%; Score 125; DB 7; Length 318;
        Best Local Similarity 100.0%; Pred. No. 6e-13;
        Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY
OY      1 SVPRHHTGVDSLCNFSPRLARRV 23
        |||||
Db      191 SVPRHHTGVDSLCNFSPRLARRV 213
ID      AAE12631 standard; protein; 367 AA.
AC      AAE12631;
XX
XX
DT      03-JAN-2002 (first entry)
DE
DE      Human gene 1 encoded secreted protein fragment, SEQ ID NO: 27.
XX
XX      Human: ovarian cancer antigen; proliferative disorder; cancer; tumour;
KW      respiratory system disorder; asthma; haemotopoietic disorder; skin aging;
KW      immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
KW      rheumatoid arthritis; inflammation; neurological disorder; septic shock;
KW      Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
KW      atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
KW      epithelial cell proliferation; transplantation; chemotaxis; infection;

```

KM food additive; wound healing; endocrine disorder; kidney disorder;
 KW gene therapy; cytostatic.
 OS Homo sapiens.
 PN WO200170804-A1.
 PD 27-SEP-2001.
 PF 16-MAR-2001; 2001WO-US008585.
 PR 17-MAR-2000; 2000US--0190076P.
 PR 23-AUG-2000; 2000US-0227009P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Birse CE, Rosen CA;
 DR WPI; 2001-639119/73.
 XX
 XX Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
 PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
 PT metastases.
 XX
 PS Disclosure; Page 9; 427pp; English.
 CC AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
 CC ovarian cancer associated protein (collectively known as ovarian cancer
 CC antigens) genes, and AAI12623-AAI12629 represent the proteins they
 CC encode. AAI12630-AAI12638 represent human ovarian cancer antigenic
 CC fragments or variants. Ovarian cancer antigens and their corresponding
 CC DNAs are used in the prevention, diagnosis and treatment of diseases
 CC associated with their inappropriate expression. These disorders include
 CC proliferative disorders, cancer, tumours, respiratory system disorders,
 CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
 CC skin disorders, autoimmune diseases (e.g., rheumatoid arthritis),
 CC inflammation, allergies, neurological disorders (e.g., Alzheimer's
 CC disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, endocrine disorders and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, to identify their cognate ligands or binding
 CC partners, in chemotaxis and can be used as a food additive. Antibodies
 CC specific for a protein of the invention can be used in alleviating
 CC symptoms associated with the disorders mentioned above and in diagnostic
 CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
 CC the invention is used in gene therapy. The present sequence represents a
 CC human ovarian cancer antigenic fragment of the invention
 CC
 SQ Sequence 367 AA;
 Query Match 100.0%; Score 125; DB 4; Length 367;
 Best Local Similarity 100.0%; Pred. No. 7,1e-13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SVPNRHTGTVDLNCNFSPLARRV 23
 ||||||||||||||||||
 DB 311 SVPNRHTGTVDLNCNFSPLARRV 333

Search completed: October 27, 2005, 18:44:25
 Job time : 80.5907 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:03:42 ; Search time 20.0047 Seconds
(without alignments)
85.826 Million cell updates/sec

Title: US-10-612-090-17
Perfect score: 125
Sequence: 1 SVPNRHTGVDSLNCNFSPLARRV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	100.0	438	4	US-09-404-879A-390 Sequence 390, App
2	125	100.0	438	4	US-09-667-857-390 Sequence 390, App
3	125	100.0	833	4	US-09-404-879A-389 Sequence 389, App
4	125	100.0	833	4	US-09-667-857-389 Sequence 389, App
5	125	100.0	914	4	US-09-404-879A-312 Sequence 312, App
6	125	100.0	914	4	US-09-338-933-312 Sequence 312, App
7	125	100.0	914	4	US-09-667-857-312 Sequence 312, App
8	51	40.8	942	3	US-09-171-461-12 Sequence 12, Appl
9	51	40.8	942	3	US-09-970-711-12 Sequence 12, Appl
10	49.5	39.6	113	4	US-09-884-570-9 Sequence 9, Appl
11	49.5	39.6	113	4	US-10-157-457A-12 Sequence 12, Appl
12	49.5	39.6	113	4	US-10-157-457A-12 Sequence 12, Appl
13	46.5	37.2	230	4	US-09-252-991A-18747 Sequence 18747, A
14	46	36.8	808	4	US-09-252-991A-19470 Sequence 19470, A
15	45	36.0	420	4	US-09-527-084A-6 Sequence 6, Appl
16	44	35.2	168	4	US-09-248-796A-14756 Sequence 14756, A
17	44	35.2	212	4	US-09-328-352-6982 Sequence 6982, Ap
18	43.5	34.8	211	4	US-09-352-991A-20225 Sequence 20225, A
19	43	34.4	80	4	US-09-248-796A-21674 Sequence 21674, A
20	43	34.4	246	4	US-09-252-991A-32673 Sequence 32673, A
21	43	34.4	410	4	US-09-552-991A-26654 Sequence 26654, A
22	43	34.4	734	4	US-09-894-998A-26 Sequence 26, Appl
23	43	34.4	734	4	US-10-337-551-26 Sequence 26, Appl
24	42.5	34.0	22	3	US-08-470-106-30 Sequence 30, Appl
25	42.5	34.0	22	3	US-08-318-794-30 Sequence 30, Appl
26	42.5	34.0	427	4	US-09-489-039A-10737 Sequence 10737, A
27	42	33.6	419	4	US-09-489-039A-13188 Sequence 13188, A

28	42	33.6	843	4	US-09-489-039A-7545 Sequence 7545, Ap
29	41.5	33.2	391	4	US-09-328-352-4861 Sequence 4861, Ap
30	41.5	33.2	1025	4	US-09-834-309-5 Sequence 5, Appl
31	41.5	33.2	1151	4	US-09-177-165A-31 Sequence 31, Appl
32	41	32.8	320	4	US-10-014-268-30 Sequence 30, Appl
33	41	32.8	321	4	US-07-945-288-6 Sequence 6, Appl
34	41	32.8	321	1	US-08-462-831-6 Sequence 6, Appl
35	41	32.8	321	1	US-08-461-809-6 Sequence 6, Appl
36	41	32.8	321	1	US-08-461-441-6 Sequence 6, Appl
37	41	32.8	321	2	US-08-482-142-6 Sequence 6, Appl
38	41	32.8	321	2	US-08-478-572-6 Sequence 6, Appl
39	41	32.8	321	3	US-08-484-286-6 Sequence 6, Appl
40	41	32.8	321	5	PCT-US93-08518-6 Sequence 6, Appl
41	41	32.8	395	4	US-09-252-991A-30497 Sequence 30497, A
42	41	32.8	490	3	US-09-099-041A-26 Sequence 26, Appl
43	41	32.8	490	3	US-09-245-281-26 Sequence 26, Appl
44	41	32.8	490	4	US-09-207-359B-26 Sequence 26, Appl
45	41	32.8	490	4	US-09-340-620A-26 Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-404-879A-390
; Sequence 390, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: ALGATE, PAUL A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-390

Query Match 100.0%; Score 125; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHTGVDSLNCNFSPLARRV 23
DB 311 SVPNRHTGVDSLNCNFSPLARRV 333

RESULT 2
US-09-667-857-390
; Sequence 390, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedyck, Thomas S.
; TITLE OF INVENTION: CARTEER, DARRICK
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-667-857-390

```

```

Query Match          100.0%; Score 125; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 6.6e-12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB      311 SVPNRHHTGVDSLNCNFSPLARRV 333

```

```

RESULT 3
US-09-404-879A-389
; Sequence 389, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-389

```

```

Query Match          100.0%; Score 125; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB      706 SVPNRHHTGVDSLNCNFSPLARRV 728

```

```

RESULT 4
US-09-667-857-389
; Sequence 389, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-667-857-389

```

```

Query Match          100.0%; Score 125; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB      706 SVPNRHHTGVDSLNCNFSPLARRV 728

```

```

RESULT 5
US-09-404-879A-312
; Sequence 312, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-404-879A-312

```

```

Query Match          100.0%; Score 125; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB      787 SVPNRHHTGVDSLNCNFSPLARRV 809

```

```

RESULT 6
US-09-338-933-312
; Sequence 312, Application US/09338933
; Patent No. 648931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-338-933-312

```

```

Query Match          100.0%; Score 125; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB      787 SVPNRHHTGVDSLNCNFSPLARRV 809

```

```

RESULT 7
US-09-667-857-312
; Sequence 312, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:

```

APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
APPLICANT: Algate, Paul A.
APPLICANT: Fling, Steven P.
APPLICANT: Retter, Marc W.
APPLICANT: Ronger, Gary Richard
APPLICANT: Reed, Steven G.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C5
CURRENT APPLICATION NUMBER: US/09/667,857
CURRENT FILING DATE: 2000-09-20
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 312
LENGTH: 914
TYPE: PRT
ORGANISM: Homo sapien
US-09-667-857-312

Query Match 100.0%; Score 125; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNRHHTGVDSLGNFSPPLARV 23
DB 787 SVNRHHTGVDSLGNFSPPLARV 809

RESULT 8
US-09-171-461-12
Sequence 12, Application US/09171461
Patent No. 6335016
GENERAL INFORMATION:
APPLICANT: Baker, Adam
APPLICANT: Cotten, Matthew
APPLICANT: Chiocca, Susanna
APPLICANT: Kurzbauer, Robert
APPLICANT: Schaffner, Gotthold
TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
FILE REFERENCE: 0652.1800000
CURRENT APPLICATION NUMBER: US/09/171,461
CURRENT FILING DATE: 1999-01-12
EARLIER APPLICATION NUMBER: PCT/EP97/01944
EARLIER FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 12
LENGTH: 942
TYPE: PRT
ORGANISM: CELO Virus
FEATURE:
OTHER INFORMATION: Position: 18289..21117 /gene: L3 /product: L3 hexon
US-09-171-461-12

Query Match 40.8%; Score 51; DB 3; Length 942;
Best Local Similarity 45.0%; Pred. No. 11;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLGNFSPPLAR 22
DB 757 PDRHYFHYDFLRNFDPMRSKQ 776

RESULT 9
US-09-970-711-12
Sequence 12, Application US/09970711
Patent No. 6773709
GENERAL INFORMATION:
APPLICANT: Baker, Adam
APPLICANT: Cotten, Matthew

APPLICANT: Chiocca, Susanna
APPLICANT: Kurzbauer, Robert
APPLICANT: Schaffner, Gotthold
TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
FILE REFERENCE: 0652.1800001
CURRENT APPLICATION NUMBER: US/09/970,711
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/171,461
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: PCT/EP97/01944
PRIOR FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 12
LENGTH: 942
TYPE: PRT
ORGANISM: CELO Virus
FEATURE:
OTHER INFORMATION: Position: 18289..21117 /gene: L3 /product: L3 hexon
US-09-970-711-12

Query Match 40.8%; Score 51; DB 4; Length 942;
Best Local Similarity 45.0%; Pred. No. 11;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLGNFSPPLAR 22
DB 757 PDRHYFHYDFLRNFDPMRSKQ 776

RESULT 10
US-09-884-570-9
Sequence 9, Application US/09884570
Patent No. 6566333
GENERAL INFORMATION:
APPLICANT: Williams, John P.
APPLICANT: McDonald, Jay M.
APPLICANT: McKenna, Margaret A.
TITLE OF INVENTION: Osteoclast Secreted Chemokine and Uses Thereof
FILE REFERENCE: D6294
CURRENT APPLICATION NUMBER: US/09/884,570
CURRENT FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/212,271
PRIOR FILING DATE: 2000-06-19
NUMBER OF SEQ ID NOS: 11
SEQ ID NO 9
LENGTH: 113
TYPE: PRT
ORGANISM: unknown
FEATURE:
LOCATION: 38..165
OTHER INFORMATION: amino acid sequence of region of Lect2
OTHER INFORMATION: protein homologous to repeat sequences of N
OTHER INFORMATION: and C terminal ends of the mim-1 protein
US-09-884-570-9

Query Match 39.6%; Score 49.5; DB 4; Length 113;
Best Local Similarity 37.5%; Pred. No. 1.8;
Matches 9; Conservative 5; Mismatches 5; Indels 5; Gaps 1;

QY 5 RHHTGVDSLGN-----FSPPLARV 23
DB 13 RHHPGVDSLGNDSGVVYAPFTGKI 36

RESULT 11
US-10-157-457A-9
Sequence 9, Application US/10157457A
Patent No. 6800610
GENERAL INFORMATION:
APPLICANT: Williams, John P.
APPLICANT: McDonald, Jay M.
APPLICANT: McKenna, Margaret A.

ORGANISM: Zea mays
;
ITS-09-527-084A-6

Query Match 36.0%; Score 45; DB 4; Length 420;
 Best Local Similarity 45.0%; Pred. No. 41;
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLNCNFSPLARR 22
 |||:|
 DB 397 PNGGSTLDPAAFSPISKR 416

Search completed: October 27, 2005, 17:34:34
 Job time : 21.0047 secs

THE NEW YORK PUBLIC LIBRARY
ASTOR LENOX TILDEN FOUNDATION
125 WEST 47TH STREET
NEW YORK, N. Y. 10036

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:30:08 ; Search time 73.6 Seconds
(without alignments)
130.638 Million cell updates/sec

Title: US-10-612-090-17
Perfect score: 125
Sequence: 1 SVPNRHHTGVDSLNCNFSPLARRV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1865214 seqs, 418043040 residues

Total number of hits satisfying chosen parameters: 1865214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10A_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUBCOMB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125	100.0	23	US-10-612-090-17	Sequence 17, Appl
2	125	100.0	108	US-10-612-090-2	Sequence 2, Appl
3	125	100.0	108	US-10-612-090-13	Sequence 13, Appl
4	125	100.0	174	US-10-383-368-4	Sequence 4, Appl
5	125	100.0	178	US-09-884-441-489	Sequence 489, App
6	125	100.0	178	US-09-907-969-489	Sequence 489, App
7	125	100.0	178	US-10-198-053-489	Sequence 489, App
8	125	100.0	178	US-10-860-790-489	Sequence 489, App
9	125	100.0	233	US-09-884-441-488	Sequence 488, App
10	125	100.0	233	US-09-907-969-488	Sequence 488, App
11	125	100.0	233	US-10-198-053-488	Sequence 488, App

12	125	100.0	233	17	US-10-860-790-488	Sequence 488, App
13	125	100.0	284	10	US-09-965-738-300	Sequence 300, App
14	125	100.0	318	10	US-09-907-969-594	Sequence 594, App
15	125	100.0	318	14	US-10-198-053-594	Sequence 594, App
16	125	100.0	318	17	US-10-860-790-594	Sequence 594, App
17	125	100.0	367	15	US-10-333-900-27	Sequence 27, Appl
18	125	100.0	396	18	US-10-858-412-225	Sequence 225, App
19	125	100.0	438	9	US-09-884-441-390	Sequence 390, App
20	125	100.0	438	9	US-09-884-441-483	Sequence 483, App
21	125	100.0	438	10	US-09-907-969-390	Sequence 390, App
22	125	100.0	438	10	US-09-907-969-483	Sequence 483, App
23	125	100.0	438	10	US-09-827-271-390	Sequence 390, App
24	125	100.0	438	14	US-10-198-053-390	Sequence 390, App
25	125	100.0	438	14	US-10-198-053-483	Sequence 483, App
26	125	100.0	438	15	US-10-333-900-19	Sequence 19, Appl
27	125	100.0	438	17	US-10-860-790-390	Sequence 390, App
28	125	100.0	438	17	US-10-860-790-483	Sequence 483, App
29	125	100.0	439	10	US-09-965-738-148	Sequence 148, App
30	125	100.0	456	18	US-10-858-412-226	Sequence 226, App
31	125	100.0	526	15	US-10-333-900-30	Sequence 30, Appl
32	125	100.0	545	14	US-10-243-243A-4	Sequence 4, Appl
33	125	100.0	583	14	US-10-142-515-4	Sequence 4, Appl
34	125	100.0	748	17	US-10-687-035-1	Sequence 1, Appl
35	125	100.0	809	17	US-10-687-035-2	Sequence 2, Appl
36	125	100.0	833	9	US-09-884-441-389	Sequence 389, App
37	125	100.0	833	10	US-09-907-969-389	Sequence 389, App
38	125	100.0	833	10	US-09-827-271-389	Sequence 389, App
39	125	100.0	833	14	US-10-198-053-389	Sequence 389, App
40	125	100.0	833	17	US-10-860-790-389	Sequence 389, App
41	125	100.0	914	9	US-09-778-320-206	Sequence 206, App
42	125	100.0	914	9	US-09-910-689-206	Sequence 206, App
43	125	100.0	914	9	US-09-884-441-312	Sequence 312, App
44	125	100.0	914	9	US-09-884-441-478	Sequence 478, App
45	125	100.0	914	10	US-09-907-969-312	Sequence 312, App

ALIGNMENTS

RESULT 1
US-10-612-090-17
; Sequence 17, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-612-090-17

Query Match 100.0%; Score 125; DB 15; Length 23;
Best Local Similarity 100.0%; Pred. No. 4,3-12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSLNCNFSPLARRV 23
DB 1 SVPNRHHTGVDSLNCNFSPLARRV 23

RESULT 2
US-10-612-090-2
; Sequence 2, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:

```

; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-612-090-2

Query Match          100.0%; Score 125; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 2,2e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SVPNRHHTGVDSLNCFSPLARRV 23
Db      37 SVPNRHHTGVDSLNCFSPLARRV 59

RESULT 3
US-10-612-090-13
; Sequence 13, Application US/10612090
; Publication No. US20040057952A1
; GENERAL INFORMATION:
; APPLICANT: Immunogen, Inc.
; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
; FILE REFERENCE: A8340
; CURRENT APPLICATION NUMBER: US/10/612,090
; CURRENT FILING DATE: 2003-07-03
; PRIOR APPLICATION NUMBER: US 60/393,094
; PRIOR FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion protein
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Glutathione S-transferase fusion site
US-10-612-090-13

Query Match          100.0%; Score 125; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 2,2e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SVPNRHHTGVDSLNCFSPLARRV 23
Db      37 SVPNRHHTGVDSLNCFSPLARRV 59

RESULT 4
US-10-383-368-4
; Sequence 4, Application US/10383368
; Publication No. US20040002091A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul E.
; APPLICANT: Ebner, Reinhard
; APPLICANT: Strovel, Jeffrey W.
; TITLE OF INVENTION: Cancer-Linked Gene as Target for Chemotherapy
; FILE REFERENCE: 689290-123
; CURRENT APPLICATION NUMBER: US/10/383,368
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US 60/362,527
; PRIOR FILING DATE: 2002-03-07

```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-383-368-4

Query Match          100.0%; Score 125; DB 15; Length 174;
Best Local Similarity 100.0%; Pred. No. 3,6e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SVPNRHHTGVDSLNCFSPLARRV 23
Db      47 SVPNRHHTGVDSLNCFSPLARRV 69

RESULT 5
US-09-884-441-489
; Sequence 489, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-489

Query Match          100.0%; Score 125; DB 9; Length 178;
Best Local Similarity 100.0%; Pred. No. 3,7e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 SVPNRHHTGVDSLNCFSPLARRV 23
Db      106 SVPNRHHTGVDSLNCFSPLARRV 128

RESULT 6
US-09-907-969-489
; Sequence 489, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Reltter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-907-969-489

Query Match 100.0%; Score 125; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.7e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSL CNFSPPLARRV 23
DB 106 SVPNRHHTGVDSL CNFSPPLARRV 128

RESULT 7

US-10-198-053-489
; Sequence 489, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198.053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-489

Query Match 100.0%; Score 125; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.7e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSL CNFSPPLARRV 23
DB 106 SVPNRHHTGVDSL CNFSPPLARRV 128

RESULT 8

US-10-860-790-489
; Sequence 489, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860.790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-489

Query Match 100.0%; Score 125; DB 17; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.7e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSL CNFSPPLARRV 23
DB 106 SVPNRHHTGVDSL CNFSPPLARRV 128

RESULT 9

US-09-884-441-488
; Sequence 488, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884.441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-488

Query Match 100.0%; Score 125; DB 9; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSL CNFSPPLARRV 23
DB 106 SVPNRHHTGVDSL CNFSPPLARRV 128

RESULT 10

US-09-907-969-488
; Sequence 488, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907.969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-969-488

Query Match 100.0%; Score 125; DB 10; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVPNRHHTGVDSL CNFSPPLARRV 23
DB 106 SVPNRHHTGVDSL CNFSPPLARRV 128

RESULT 11

US-10-198-053-488
; Sequence 488, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-488

Query Match          100.0%; Score 125; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVPNRHHGTGVDSLCNFSPFLARRV 23
Db 106 SVPNRHHGTGVDSLCNFSPFLARRV 128

RESULT 12
US-10-860-790-488
; Sequence 488, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-488

Query Match          100.0%; Score 125; DB 17; Length 233;
Best Local Similarity 100.0%; Pred. No. 4.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVPNRHHGTGVDSLCNFSPFLARRV 23
Db 106 SVPNRHHGTGVDSLCNFSPFLARRV 128

RESULT 13
US-09-965-738-300
; Sequence 300, Application US/09965738
; Publication No. US20030143667A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy
; TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic a
; FILE REFERENCE: 40715-258841
; CURRENT APPLICATION NUMBER: US/09/965,738
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/284,175
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 300
; LENGTH: 284

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-738-300

Query Match          100.0%; Score 125; DB 10; Length 284;
Best Local Similarity 100.0%; Pred. No. 6e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVPNRHHGTGVDSLCNFSPFLARRV 23
Db 157 SVPNRHHGTGVDSLCNFSPFLARRV 179

RESULT 14
US-09-907-969-594
; Sequence 594, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Flinn, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriack
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-09-907-969-594

Query Match          100.0%; Score 125; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 6.8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SVPNRHHGTGVDSLCNFSPFLARRV 23
Db 191 SVPNRHHGTGVDSLCNFSPFLARRV 213

RESULT 15
US-10-198-053-594
; Sequence 594, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594
; LENGTH: 318

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-10-198-053-594

```

```

Query Match          100.0%; Score 125; DB 14; Length 318;
Best Local Similarity 100.0%; Pred No. 6,8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  SYPNRHHTGVDSLGNFSPPLARRV 23
          |||||
Db      191 SYPNRHHTGVDSLGNFSPPLARRV 213

```

Search completed: October 27, 2005, 18:31:47
 Job time : 74.6 secs

Tim Pegg Book (1970)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 18:13:49 ; Search time 14.2279 Seconds
(without alignments)
155.538 Million cell updates/sec

Title: US-10-612-090-17
Perfect score: 125
Sequence: 1 SVPNRHHTGVDSLNCNFSPLARRV 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	41.6	390	2 F84196	hypothetical prote
2	49	39.2	209	2 T44653	capsular polysacch
3	49	39.2	266	2 T15066	hypothetical prote
4	48	38.4	462	2 T25726	hypothetical prote
5	47	37.6	294	2 AF1218	Salmonella typhim
6	47	37.6	294	2 A11571	Regulatory protein
7	46	36.8	774	2 D83507	hypothetical prote
8	45	36.0	285	2 A82609	hypothetical prote
9	45	36.0	366	2 T25178	hypothetical prote
10	45	36.0	420	2 T01172	G1/S transilition co
11	43.5	34.8	332	2 C83933	penicillin-binding
12	43.5	34.8	403	2 H64861	hypothetical prote
13	43	34.4	325	2 B82201	conserved hypotet
14	43	34.4	335	2 G64817	probable membrane
15	43	34.4	344	1 WMBR31	38k protein - huma
16	43	34.4	530	2 T18596	hypothetical prote
17	43	34.4	735	1 WMBRT5	U15 protein - hum
18	43	34.4	1018	2 T19693	hypothetical prote
19	43	34.4	2160	2 T20241	hypothetical prote
20	42.5	34.0	160	2 I84444	eosinophil-derived
21	42	33.6	117	1 ERADD1	early E3 13k glyco
22	42	33.6	810	1 P2MMBB	conserved hypotet
23	42	33.6	810	1 P2MMBB	2a protein - broad
24	42	33.6	813	2 T21192	hypothetical prote
25	41.5	33.2	120	2 E70738	hypothetical prote
26	41.5	33.2	232	2 E82140	probable C4-dicard
27	41.5	33.2	508	2 S51809	phosphoprotein pho
28	41.5	33.2	591	2 H86267	probable protein p
29	41.5	33.2	982	2 T43699	DNA mismatch repai

30	41.5	33.2	1025	1 A43526	complement C3d/Bps
31	41.5	33.2	1151	2 A41529	GR1 protein - yea
32	41.5	33.2	1239	2 T06143	disease resistance
33	41.5	33.2	3944	2 T19997	hypothetical prote
34	41	32.8	152	2 F84642	hypothetical prote
35	41	32.8	168	2 T28745	hypothetical prote
36	41	32.8	280	2 T17116	protein kinase cdc
37	41	32.8	319	2 A61500	allergen Der f I p
38	41	32.8	335	2 G85589	hypothetical prote
39	41	32.8	335	2 E90739	hypothetical prote
40	41	32.8	350	2 E84109	potassium uptake p
41	41	32.8	410	2 S77844	alanine-tRNA ligase
42	41	32.8	453	2 A96688	hypothetical prote
43	41	32.8	487	2 T34868	probable glutamate
44	41	32.8	589	2 AD2263	hypothetical prote
45	41	32.8	711	2 P65201	phosphoenolpyruvat

ALIGNMENTS

RESULT 1

F84196 hypothetical protein Vng0378c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: F84196
R:Ng, W.V.; Kennedy, S.P.; Mahataas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danison, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freilich, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: F84196
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-390 <STO>
A:Cross-references: UNIPROT:Q9H570; GB:AE004437; NID:G10579994; PIDN:AA618938.1; GSPDB:G C:Genetics:
A:Gene: VNG0378C

Query Match 41.6%; Score 52; DB 2; Length 390;
Best Local Similarity 56.2%; Pred. No. 1.3;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLNCNFSPL 18
Db 41 PNAHNAVVDLSRSLVAP 56

RESULT 2

T44653 capsular polysaccharide biosynthesis protein cpsE [imported] - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T44653
R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; NittayaJarn, A.; Rubens, C.E. submitted to the EMBL Data Library, June 1999
A:Reference number: Z22821
A:Accession: T44653
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-209 <CHA>
A:Cross-references: UNIPROT:Q9RPB8; UNIPROT:Q8DZ55; EMBL:AF163833; PIDN:AA53076.1
A:Experimental source: strain COH1; serotype III
C:Genetics:
A:Gene: cpsE

Query Match 39.2%; Score 49; DB 2; Length 209;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HHTGVDSLGNFSPPLA 20
DB 138 HHTVESHCHNIAFNA 152

RESULT 3
T15066
hypothetical protein R02D3.8 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15066
R:Mu, X.; Antoniou, B.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid R02D3.
A:Reference number: Z18284
A:Accession: T15066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-266 <MIX>
A:Cross-references: UNIPROT:Q9TZ29; EMBL:AF038615; NID:g2736322; PID:g2736329; PIDN:AAB9
A:Experimental source: strain Bristol N2; clone R02D3
C:Genetics:
A:Gene: CESP:R02D3.8
A:Map position: 4
A:Introns: 17/1; 51/1; 127/3; 185/2; 220/3
C:Superfamily: *Caenorhabditis elegans* hypothetical protein M02B7.2

Query Match 39.2%; Score 49; DB 2; Length 266;
Best Local Similarity 41.2%; Pred. No. 2.6;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 5 RHHTGVDSLGNFSPPLAR 21
DB 222 RHSGIDDVANICEIVR 238

RESULT 4
T25726
hypothetical protein F25B4.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25726
R:Gertung, S.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid F25B4.
A:Reference number: Z20076
A:Accession: T25726
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-462 <GAT>
A:Cross-references: UNIPROT:P54871; EMBL:U64842; PIDN:AAB37084.1; GSPDB:GN00023; CESP:F2
A:Experimental source: strain Bristol N2; clone F25B4
R:Boukattane, Y.; Duncan, A.; Wang, S.; Labuda, D.; Robert, M.F.; Sarrazin, J.; Schapet
Genomes 23; 552-559, 1994
A:Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, c
A:Reference number: A55729; MUID:95154824; PMID:7851882
A:Accession: G5729
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 5-462 <BOU>
A:Cross-references: GB:U1287; NID:g555828; PIDN:AA92672.1; PID:g555829
C:Genetics:
A:Gene: CESP:F25B4.6
A:Map position: 5
A:Introns: 35/3; 78/3; 129/3; 205/1; 283/2; 346/2; 417/1
C:Superfamily: hydroxymethylglutaryl-CoA synthase

Query Match 38.4%; Score 48; DB 2; Length 462;
Best Local Similarity 42.3%; Pred. No. 7;
Matches 11; Conservative 3; Mismatches 6; Indels 6; Gaps 1;

OY 4 NRHTGVDSLGNF-----SPLARV 23
DB 111 NRHTGVDSLGNF-----SPLARV 23

DB 239 NRHTTGIDGLNSFDGVLHSPFTKRV 264

RESULT 5
AF1218
Salmonella typhimurium PCR protein homolog lmo1150 [imported] - *Listeria monocytogenes*
C:Species: *Listeria monocytogenes*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF1218
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurgat, O.; Entlian, K.D.; Fahl, H.;
Science 294, 849-852, 2001
A:Authors: Krefl, U.; Kunz, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.; Mat
ok, C.; Schueller, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1218
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <GLA>
A:Cross-references: UNIPROT:Q8Y7W9; GB:NC_003210; PIDN:CAC99228.1; PID:g16410566; GSPDB:C
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1150

Query Match 37.6%; Score 47; DB 2; Length 294;
Best Local Similarity 44.4%; Pred. No. 6.2;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 6 HHTGVDSLGNFSPPLARV 23
DB 33 HGTEVSRICNFTPCOLI 50

RESULT 6
A11571
Regulatory protein *Salmonella typhimurium* PCR protein homolog l1n114 [imported] - *Liste*
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: A11571
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurgat, O.; Entlian, K.D.; Fahl, H.;
Science 294, 849-852, 2001
A:Authors: Krefl, U.; Kunz, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maltournam, A.; Mat
ok, C.; Schueller, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11571
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <GLA>
A:Cross-references: UNIPROT:Q92CQ7; GB:AL592022; PIDN:CAC96345.1; PID:g16413573; GSPDB:G
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: l1n114

Query Match 37.6%; Score 47; DB 2; Length 294;
Best Local Similarity 44.4%; Pred. No. 6.2;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 6 HHTGVDSLGNFSPPLARV 23
DB 33 HGTEVSRICNFTPCOLI 50

RESULT 7
D83507
hypothetical protein PA115 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: D83507


```

A:Gene: BH2267
Query Match          34.8% Score 43.5; DB 2; Length 332;
Best Local Similarity 45.5% Pred. No. 27;
Matches 10; Conservative 4; Mismatches 5; Indels 3; Gaps 2;

OY      2 VPNRHRT--GVDSLGN-FSPILA 20
      ||| ||| | | | | | | |
Db      41 IPNEHTRFGIASGCKLFTALA 62

RESULT 12
H64861
hypothetical protein b163 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: H64861
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
A.: Rose, D.J.; Mau, B.; Shaio, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64861
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <BLAT>
A:Cross-references: UNIPROT:P75990; GB:AE000215; GB:U00096; NID:g1787405; PIDN:AACT4247
A:Experimental source: strain K-12, substrain MG1655

Query Match          34.8% Score 43.5; DB 2; Length 403;
Best Local Similarity 50.0% Pred. No. 33;
Matches 9; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY      5 RHHTGVDSLGNFSPILARR 22
      || | | | | | | | |
Db      70 RHYNIVELCDYAP-ARR 86

RESULT 13
B82301
conserved hypothetical protein VC0614 [imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: B82301
R:Heidelberg, J.F.; Sisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
charison, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82301
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-325 <HEI>
A:Cross-references: UNIPROT:Q9KUA9; GB:AE004147; GB:AE003852; NID:g9655045; PIDN:AAF937B7
A:Experimental source: serogroup O1, strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0614
A:Map position: 1

Query Match          34.4% Score 43; DB 2; Length 325;
Best Local Similarity 33.3% Pred. No. 32;
Matches 6; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY      6 HHTGVDSLGNFSPILARRV 23
      || | | | | | | | |
Db      269 HHKGAERICLMGSIARI 286

RESULT 14
G64817
probable membrane protein ybin - Escherichia coli (strain K-12)
C:Species: Escherichia coli

```

[illegible]

```

Oy      1  SVPNRHHT---GVDSL---CNFSPPL---ARRV 23
      : | | | | | | | | | | | | | | | | | |
Db      51  ATPKRHQIVPFGVGTLDHDCCEHSPLEFSAVARRL 83

Search completed: October 27, 2005, 18:57:27
Job time : 15.2279 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:34:49 ; Search time 66.8605 Seconds
(without alignments)
176.155 Million cell updates/sec

Title: US-10-612-090-17
Perfect score: 125
Sequence: 1 SVPNRHHTGVDSLCNFSPLARRV 23

Scoring table: BLOSUM62
Gapex 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	100.0	1148	2 Q9H7S7	Q9H7S7 homo sapien
2	125	100.0	6995	2 Q96RK2	Q96RK2 homo sapien
3	125	100.0	22152	2 Q8WJ17	Q8WJ17 homo sapien
4	98	78.4	258	2 Q9D1H1	Q9D1H1 mus musculu
5	52	41.6	141	2 Q96C55	Q96C55 rhizobium l
6	52	41.6	390	1 Y378 HALN1	Q9H670 halobacteri
7	51	40.8	942	1 HEX ADEG1	Q96585 avian adeno
8	51	40.8	942	2 Q96585	Q96585 avian adeno
9	51	40.8	942	2 Q7M5F4	Q96585 avian adeno
10	50.5	40.4	785	2 Q614N2	Q614N2 oryza sativ
11	49.5	39.6	151	1 LCT2 MOUSE	Q88803 mus musculu
12	49.5	39.6	151	2 Q8K181	Q8K181 mus musculu
13	49	39.2	207	2 Q9XDT0	Q9XDT0 streptococc
14	49	39.2	209	2 Q9RPB8	Q9RPB8 streptococc
15	49	39.2	209	2 Q8DZES	Q8DZES streptococc
16	49	39.2	209	2 Q8E503	Q8E503 streptococc
17	49	39.2	266	2 Q9T2Z9	Q9T2Z9 caenorhabdi
18	48.5	38.8	202	2 Q9CKQ5	Q9CKQ5 pasteurilla
19	48.5	38.8	521	2 Q6FUK9	Q6FUK9 candida gla
20	48.5	38.8	869	2 Q6GMB5	Q6GMB5 paracitlamy
21	48	38.4	462	1 HMGCS CAEEL	P54971 caenorhabdi
22	47	37.6	198	2 Q9L180	Q9L180 streptomyce
23	47	37.6	294	2 Q8Y7W9	Q8Y7W9 listeria mo
24	47	37.6	294	2 Q92CQ7	Q92CQ7 listeria in
25	47	37.6	294	2 Q720T1	Q720T1 listeria mo
26	47	37.6	541	1 FTCD HUMAN	Q95954 h formimid
27	47	37.6	554	2 Q86V03	Q86V03 homo sapien
28	46.5	37.2	1750	2 Q9RPF5	Q9RPF5 stigmatella
29	46	36.8	405	2 Q98L46	Q98L46 rhizobium l
30	46	36.8	440	2 Q8FT90	Q8FT90 corynebacte
31	46	36.8	496	2 Q93J15	Q93J15 streptomyce

32	46	36.8	774	2 Q914M0	Q914M0 pseudomonas
33	46	36.8	873	1 SYA WIGBR	Q8d3w8 wigglewort
34	46	36.8	1981	2 Q9E9R6	Q9E9R6 hop latent
35	45.5	36.4	1045	2 Q4G545	Q4G545 ovie aries
36	45.5	36.4	1343	1 TIM_DROVI	Q17482 drosophila
37	45	36.0	163	2 Q8D1B0	Q8D1B0 synchococc
38	45	36.0	242	2 Q852C9	Q852C9 oryza sativ
39	45	36.0	245	2 Q7N564	Q7N564 photorhabdu
40	45	36.0	265	1 UPPI_BACAN	Q81vc9 bacillus an
41	45	36.0	265	1 UPPI_BACHK	Q6hpb7 bacillus th
42	45	36.0	265	2 Q63G06	Q63G06 bacillus ce
43	45	36.0	285	2 Q9PB06	Q9PB06 xylella fas
44	45	36.0	356	1 YOP3 CAEEL	Q22695 caenorhabdi
45	45	36.0	420	2 Q22345	Q22345 zea mays (m

ALIGNMENTS

RESULT 1

Q9H7S7 PRELIMINARY; PRT; 1148 AA.

AC Q9H7S7; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ14303.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishii T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fuji A., Hara H., Tanase T., Nomura Y.,
RA Toguya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima N., Sano S.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiyawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.;
*Complete sequencing and characterization of 21,243 full-length human
cDNAs.*
RT Nat. Genet. 36:40-45(2004).
RL EMBL; AK024365; BAB14899.1; -.
DR HSSP; Q9D1H1; 11VZ.
DR InterPro; IPR000082; SEA.
DR pfam; PF01390; SEA; 7.
DR PROSITE; PS50024; SEA; 3.
SQ SEQUENCE 1148 AA; 127957 MW; 3861B0D5EFPD8ABC CRC64;

```

Query Match          100.0%; Score 125; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 4.2e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVPNRHHTGVDSLGNFSPPLARRV 23
Db      1021 SVPNRHHTGVDSLGNFSPPLARRV 1043

RESULT 2
Q96RK2      PRELIMINARY;      PRT; 6995 AA.
AC      O96RK2.
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Mucin 16 (Fragment).
GN      Name=MUC16;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21347965; PubMed=11369781; DOI=10.1074/jbc.M103554200;
RA      Yin B.W., Lloyd K.O.;
RT      "Molecular cloning of the cal25 ovarian cancer antigen. Identification
as a new mucin, muc16."
RL      J. Biol. Chem. 276:27371-27375(2001).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Lloyd K.O., Yin B.W.T.;
RT      Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF161486; AAK74120.3; -.
DR      HSSP; Q9D1H1; 11VZ.
DR      InterPro; IPR000194; ATPase_a/bcentre.
DR      InterPro; IPR000082; SEA.
DR      Pfam; PF01390; SEA; 20.
DR      SMART; SM00200; SEA; 10.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR      PROSITE; PS50024; SEA; 6.
RT      NON TER
SQ      SEQUENCE 6995 AA; 744958 MW; 80C797DBDF33A2B CRC64;

Query Match          100.0%; Score 125; DB 2; Length 6995;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVPNRHHTGVDSLGNFSPPLARRV 23
Db      6868 SVPNRHHTGVDSLGNFSPPLARRV 6890

RESULT 3
Q8WX17      PRELIMINARY;      PRT; 22152 AA.
AC      Q8WX17.
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Ovarian cancer related tumor marker CA125.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21646939; PubMed=11786729; DOI=10.1159/000050638;
RA      O'Brien T.J., Beard J.B., Underwood L.J., Dennis R.A., Santini A.D.,
RA      York L.;
RT      "The CA 125 gene: an extracellular superstructure dominated by repeat
sequences."
RL      Tumour Biol. 22:348-366(2001).

```

```

RN      [2]
RP      SEQUENCE FROM N.A.
RA      O'Brien T.J., Underwood L.J., Beard J.B.;
RT      Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF144442; AA65133.2; -.
DR      Genew; HGNC:15582; MUC16.
DR      Pfam; PF01390; SEA; 51.
DR      SMART; SM00200; SEA; 23.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_2.
DR      PROSITE; PS50024; SEA; 11.
SQ      SEQUENCE 22152 AA; 2352668 MW; B3E7BDF1997A440 CRC64;

Query Match          100.0%; Score 125; DB 2; Length 22152;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SVPNRHHTGVDSLGNFSPPLARRV 23
Db      22025 SVPNRHHTGVDSLGNFSPPLARRV 22047

RESULT 4
Q9D1H1      PRELIMINARY;      PRT; 258 AA.
AC      Q9D1H1.
DT      01-JUN-2001 (TrEMBLrel. 17, Created)
DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched
library, clone:1110008114 product:hypothetical SEA domain containing
protein, full insert sequence.
GN      Name=1110008114Rik;
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=whole body;
RX      MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA      Carninci P., Hayashizaki Y.;
RT      "High-efficiency full-length cDNA cloning."
RL      Meth. Enzymol. 303:19-44(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=whole body;
RX      MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA      RIKEN FANTOM Consortium;
RT      "Functional annotation of a full-length mouse cDNA collection."
RL      Nature 409:685-690(2001).
RN      [3]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=whole body;
RT      The FANTOM Consortium,
RA      the RIKEN Genome Exploration Research Group Phase I & II Team;
RT      "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL      Nature 420:563-573(2002).
RN      [4]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=whole body;
RX      MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA      Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA      Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT      "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes."
RL      Genome Res. 10:1617-1630(2000).
RN      [5]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=whole body;
RX      MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA      Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA      Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,

```

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayaishi Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-Format
 sequencing pipeline with 384 multicapillary sequencer.",
 RL Genome Res. 10:1757-1771(2000).
 [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=whole body;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carriacci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hasegaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK003577; BAB22869.1; -.
 DR PDB: 1IVZ; NMR: A=60-190.
 DR MGI: 1320982; 111008114RIK.
 DR InterPro: IPR000082; SEA.
 DR Pfam: PF01390; SEA: 1.
 DR PROSITE: PS50024; SEA: 1.
 DR Hypothetical protein.
 KW SEQUENCE 258 AA; 29425 MW; B64D9B63394D84E7 CRC64;
 SQ
 Query Match 78.4%; Score 98; DB 2; Length 258;
 Best Local Similarity 90.0%; Pred. No. 1.9e-07; Indels 0; Gaps 0;
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 NRHHGTGVDSLCNFSPPLARRV 23
 Db 134 NNNHTGVDSLCNFSPPLARRV 153
 RESULT 5
 Q98C55 PRELIMINARY; PRT; 141 AA.
 AC Q98C55;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE M115290 protein.
 GN OrderedLocNames=M115290;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.",
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003006; BAB51766.1; -.
 KW Complete proteome.
 SQ SEQUENCE 141 AA; 15437 MW; C43AC343668BEDAD CRC64;
 Query Match 41.6%; Score 52; DB 2; Length 141;
 Best Local Similarity 40.0%; Pred. No. 2.9; Mismatches 8; Conservative 6; Indels 0; Gaps 0;

QY 2 VPRHHGTGVDSLCNFSPPLAR 21
 Db 59 IPSKLPALFVCHFSRPAER 78
 RESULT 6
 Y378 HALNI STANDARD; PRT; 390 AA.
 ID Y378 HALNI
 AC Q9H570;
 DT 28-FEB-2003 (rel. 41, Created)
 DT 28-FEB-2003 (rel. 41, Last sequence update)
 DT 05-JUL-2004 (rel. 44, Last annotation update)
 DE Hypothetical UPF0204 protein Vng0378C.
 GN OrderedLocNames=VNG0378C;
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20504483; PubMed=11016950; DOI=10.1073/pnas.190337797;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Geo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddock D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Beck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis F.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RT "Genome sequence of Halobacterium species NRC-1.",
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -1-SIMILARITY: Belongs to the UPF0204 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AE004995; AAG18938.1; -.
 DR PIR: P84196; F84196.
 DR HAMAP: MF_00562; atypical; 1.
 DR InterPro: IPR007508; DUF516.
 DR Pfam: PF04414; DUF516; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 390 AA; 40610 MW; A2AF934FF59498317 CRC64;
 Query Match 41.6%; Score 52; DB 1; Length 390;
 Best Local Similarity 56.2%; Pred. No. 8.7; Indels 0; Gaps 0;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 3 PNRHHGTGVDSLCNFSP 18
 Db 41 PNAHNAVVDLSRVP 56
 RESULT 7
 HEX_ADEG1 STANDARD; PRT; 942 AA.
 ID HEX_ADEG1
 AC P42511; O64758;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 05-JUL-2004 (rel. 44, Last annotation update)
 DE Hexon protein (Late protein 2).
 GN Name=PII;
 OS Avian adenovirus gall (strain Phelps) (Powl adenovirus 1) (CELO).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenoviruses.
 OX NCBI_TaxID=10553;
 RN [1]
 RP SEQUENCE FROM N.A.


```

RX MEDLINE=96186720; PubMed=8627769;
RA Chlocca S., Kurzhauser R., Schaffner G., Baker A., Mautner V.,
RA Cocten M.;
RT "The complete DNA sequence and genomic organization of the avian
RT Adenovirus CEO.";
RL J. Virol. 70:2939-2949(1996).
RN
RP SEQUENCE OF 913-942 FROM N.A.
RX MEDLINE=93362429; PubMed=8395124;
RA Cai F., Weber J.M.;
RT "Organization of the avian adenovirus genome and the structure of its
RT endopeptidase.";
RL Virology 196:358-362(1993).
CC -1 FUNCTION: This protein is one of the structural proteins in the
CC viral coat and is synthesized during late infection (By
CC similarity).
CC -1 SUBUNIT: Homotrimer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U46933; AAC54912.1; -.
DR EMBL: L13161; AAA51401.1; -.
DR HSSP: P03277; IDHX.
DR InterPro: IPR000736; Adeno_hexon.
DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon_C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
DR Coar protein; Hexon protein; Late protein.
SQ SEQUENCE 942 AA; 106709 MW; 7F4CE8D3F17D051B CRC64;

Query Match 40.8%; Score 51; DB 1; Length 942;
Best Local Similarity 45.0%; Pred. No. 33;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLGNFSPPLARR 22
Db 757 PDRHYFHYDFLRNFDPMRSRQ 776

RESULT 8
Q96585 PRELIMINARY; PRT; 942 AA.
AC Q96585;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Viral structural protein.
GN Name=hexon;
OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10553;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=FAV1;
RX MEDLINE=97049067; PubMed=8893797;
RA Akopian T.A., Doronin K.K., Karpov V.A., Naroditsky B.S.;
RT "Sequence of the avian adenovirus FAV1 (CELO) DNA encoding the hexon-
RT associated protein pIV and hexon.";
RL Arch. Virol. 141:1759-1765(1996).
DR EMBL: Z67970; CAA91908.1; -.
DR HSSP: P04133; IP30.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR000736; Adeno_hexon.
DR InterPro: IPR001680; WD40.
DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon_C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
DR PROSITE: PS00678; WD REPEATS_1; UNKNOWN 1.
SQ SEQUENCE 942 AA; 106709 MW; 7F4CE8D3F17D051B CRC64;

```

```

DR ProDom: PD002815; Adeno_hexon; 1.
DR PROSITE: PS00678; WD REPEATS_1; UNKNOWN 1.
SQ SEQUENCE 942 AA; 106695 MW; 903732D3F5E6DC6F CRC64;

Query Match 40.8%; Score 51; DB 2; Length 942;
Best Local Similarity 45.0%; Pred. No. 33;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLGNFSPPLARR 22
Db 757 PDRHYFHYDFLRNFDPMRSRQ 776

RESULT 9
Q7MSF4 PRELIMINARY; PRT; 942 AA.
ID Q7MSF4;
AC Q7MSF4;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE Hexon.
OS Avian adenovirus gall (strain Phelps) (Fowl adenovirus 1) (CELO).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=10553;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22935380; PubMed=14573794; DOI=10.1099/vir.0.19497-0;
RA Davidson A.J., Benko M., Harrach B.;
RT "Genetic content and evolution of adenoviruses.";
RL J. Gen. Virol. 84:2895-2908(2003).
CC -1 MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/genbank/DBJ third party annotation (TPA) entry.
DR EMBL: BK001452; DAA01617.1; -.
DR HSSP: P04133; IP30.
DR GO: GO:0019028; F:structural molecule activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR000736; Adeno_hexon.
DR InterPro: IPR001680; WD40.
DR Pfam: PF01065; Adeno_hexon; 1.
DR Pfam: PF03678; Adeno_hexon_C; 1.
DR ProDom: PD002815; Adeno_hexon; 1.
DR PROSITE: PS00678; WD REPEATS_1; UNKNOWN 1.
SQ SEQUENCE 942 AA; 106709 MW; 7F4CE8D3F17D051B CRC64;

Query Match 40.8%; Score 51; DB 2; Length 942;
Best Local Similarity 45.0%; Pred. No. 33;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNRHHTGVDSLGNFSPPLARR 22
Db 757 PDRHYFHYDFLRNFDPMRSRQ 776

RESULT 10
Q6LAN2 PRELIMINARY; PRT; 785 AA.
ID Q6LAN2;
AC Q6LAN2;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein P0473H02.11.
GN Name=P0473H02.11;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC Ehrhacoidae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN
RP SEQUENCE FROM N.A.
RA Chow T.-Y., Heng Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
RA Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,
RA Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Lwu H.-L.,

```



```

RESULT 11
LCRT2_MOUSE          STANDARD;          PRT;          151 AA.
ID LCRT2_MOUSE
O88803; O88804; Q9QWN3; Q9Z337;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Leukocyte cell-derived chemotaxin 2 precursor (Chondromodulin II)
   (Cm-II).
GN Name=Lect2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
   [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT VAL-129.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=98385586; PubMed=9714793; DOI=10.1016/S0378-1119(98)00294-7;
RA Yamagoe S., Watanabe T., Mizuno S., Suzuki K.;
RT "The mouse Lect2 gene: Cloning of cDNA and genomic DNA, structural
RL characterization and chromosomal localization."
   Gene 216:171-178(1998).
   [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Swiss Webster / NIH; TISSUE=Embryo, and Liver;
RX MEDLINE=99160594; PubMed=10050029;
RA Shukunami C., Kondo J., Wakai H., Takahashi K., Inoue H., Kamizono A.,
RA Hiraki Y.;
RT "Molecular cloning of mouse and bovine chondromodulin-II cDNAs and the
RT growth-promoting actions of bovine recombinant protein."
RL J. Biochem. 125:436-442(1999).
CC -1- FUNCTION: Has a neutrophil chemotactic activity. Also a positive
CC regulator of chondrocyte proliferation.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Lect2;
CC IsoId=O88803-1; Sequence=Displayed;
CC Name=2; Synonyms=Lect20;
CC IsoId=O88803-2; Sequence=VSP_003051;
CC -1- TISSUE SPECIFICITY: Highly expressed in liver and weakly in
CC testis. Not expressed in heart, brain, spleen, lung, skeletal
CC muscle and kidney.
CC -1- SIMILARITY: Belongs to the LECT2 / MIM-1 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL; AB009687; BAA3383.1; -.

```

DR	EMBL; AB009688; BAA33384.1; -
DR	EMBL; AB009689; BAA33385.1; -
DR	EMBL; AB009689; BAA33386.1; -
DR	EMBL; AF035161; AAF31302.1; -
DR	MED; MGI:1278342; Lect2.
DR	InterPro; IPR008663; LECT2.
DR	Pfam; PF05429; LECT2; 1.
KM	Alternative splicing; Chemotaxis; Signal.
FT	SIGNAL
FT	CHAIN
FT	VARSPLIC
FT	FT
FT	FT
SO	VARIANT
SO	SEQUENCE
Qy	5 RHHTGVDSLNC-----FSPLARRV 23
Dd	51 RHHGVDVLCSDGSVYAPFTGKI 74
RESULT 12	
ID	08K181 PRELIMINARY; PRT; 151 AA.
AC	08K181;
DT	01-OCT-2002 (TrEMBLrel. 22, Created)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
GN	Leukocyte cell-derived chemotaxin 2.
DE	Name=Lect2;
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FVB/N; TISSUE=Liver;
RX	MEDLINE=22386257; PubMed=12477937; DOI=10.1073/pnas.242603899;
RA	Straubeberg R.L., Feltingold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shamen C.M., Schuler G.D.,
RA	Aleschul S.F., Zeeberg B., Butow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marzinska K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein W.J., Ueda T.B., Toshitsuki S., Carimaci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA	Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulky S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield V.S.,
RA	Krzywnicki M.I., Skalecki U., Smallie D.E., Schercher A., Schein J.E.,
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
RN	[2]
RN	SEQUENCE FROM N.A.
RC	STRAIN=FVB/N; TISSUE=Liver;
RA	Straubeberg R.;
DR	Submitted (Apr-2002) to the EMBL/GenBank/DDBJ databases.
DR	MED; BC027753; AAH27753.1; -
DR	MED; MGI:1278342; Lect2.
DR	GO; GO:0005615; C:extracellular space; TAS.
DR	InterPro; IPR008663; LECT2.
DR	Pfam; PF05429; LECT2; 1.

```

SQ SEQUENCE 151 AA; 16439 MW; B7AF445C2B017FE2 CRC64;
Query Match 39.6%; Score 49.5; DB 2; Length 151;
Best Local Similarity 37.5%; Pred. No. 7.8;
Matches 9; Conservative 5; Mismatches 5; Indels 5; Gaps 1;

OY 5 RHHTGVDSLNCN----FSPPLARV 23
   ||| ||| ||| : : : : :
   ||| ||| ||| : : : : :
Db 51 RHHPGVDLCSGDSVYVAFPTGKI 74

RESULT 13
O9XDT0 PRELIMINARY; PRT; 207 AA.
AC O9XDT0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Neud.
GN Name=neud;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395021; PubMed=10464185;
RA Yamamoto S., Miyake K., Koike Y., Watanabe M., Machida Y., Ohta M.,
RA Iijima S.;
RA "Molecular characterization of type-specific capsular polysaccharide
RT biosynthesis genes of Streptococcus agalactiae type Ia.";
RL J. Bacteriol. 181:5176-5184(1999).
DR EMBL; AB028896; BAA82289.2; -
DR EMBL; IPRO11004; Trimer LpxA like.
SQ SEQUENCE 207 AA; 22397 MW; D3329979F9E68C53 CRC64;

Query Match 39.2%; Score 49; DB 2; Length 207;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HHTGVDSLNCNFSPLA 20
   ||| ||| ||| : : : : :
   ||| ||| ||| : : : : :
Db 138 HHTVESHCHNIAFNA 152

RESULT 14
O9RPB8 PRELIMINARY; PRT; 209 AA.
AC O9RPB8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Neud.
GN Name=neud;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COH1;
RC MEDLINE=96228704; PubMed=8830246;
RA Haft R.F., Weesels M.R., Mebane M.F., Conaty N., Rubens C.E.;
RA "Characterization of cpsF and its product CMP-N-acetylmuramic acid
RT synthetase, a group B streptococcal enzyme that can function in KI
RT capsular polysaccharide biosynthesis in Escherichia coli.";
RL Mol. Microbiol. 19:555-563(1996).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=COH1;
RX DOI=10.1128/DB.182.16.4466-4477.2000;
RA Chaffin D.O., Beres S.B., Yim H.H., Rubens C.E.;

```

```

RT "The serotype of type Ia and III group B streptococci is determined by
RT the polymerase gene within the polyclonistic capsule operon.";
RL J. Bacteriol. 182:4466-4477(2000).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=NT6;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=CNCTC 1/82;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=CNCTC 1/82;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RC STRAIN=SMU014, and 7271;
RA Cieslewicz M.J., Glusman G., Chaffin D., Kasper D., Madan A.,
RA Rodrigues S., Fahey J., Weesels M.R., Rubens C.E.;
RT "Evolution of group B Streptococcus capsular polysaccharides.";
RT Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF163833; AAD53076.1; -
DR EMBL; AF163833; AAD53076.1; -
DR EMBL; AF349539; AAK29663.1; -
DR EMBL; AF355776; AAK3617.1; -
DR EMBL; AY375362; AAR29918.1; -
DR EMBL; AY375363; AAR29920.1; -
DR EMBL; AY376403; AAR29958.1; -
DR PIR; T44653; T44653.
DR InterPro; IPRO11004; Trimer LpxA like.
SQ SEQUENCE 209 AA; 22578 MW; B007433783C852E3 CRC64;

Query Match 39.2%; Score 49; DB 2; Length 209;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 HHTGVDSLNCNFSPLA 20
   ||| ||| ||| : : : : :
   ||| ||| ||| : : : : :
Db 138 HHTVESHCHNIAFNA 152

RESULT 15
O8DZES PRELIMINARY; PRT; 209 AA.
AC O8DZES;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Neud protein.
GN Name=neud; OrderedLocuNames=SA61159;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RC MEDLINE=22229988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tetteelin H., Maignani V., Cieslewicz M.J., Eileen J.A., Peterson S.N.,
RA Weesels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA Debay R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedotova N.B., Scandlen D., Knout H.M., Mulligan S.,
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo P., Rapaport R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
 DR EMBL; AE014244; AAN00041.1; -.
 DR PIR; T44653; T44653.
 DR TIGR; SAG1159; -.
 DR InterPro; IPR011004; Trimer_LpxA_like.
 KW Complete proteome.
 SQ SEQUENCE 209 AA; 22578 MW; B007433783C852E3 CRC64;

Query Match 39.2%; Score 49; DB 2; Length 209;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 6 HHTGVDSLGNFSPPLA 20
 |||:|:|:|:
 Db 138 HHTVESHCHNAPNA 152

Search completed: October 27, 2005, 18:55:21
 Job time : 69.8605 secs

١٠٠

Query Match 100.0%; Score 141; DB 8; Length 28;
 Best Local Similarity 100.0%; Pred. No. 7.3e-15;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DRVAIYEEFLMRTNGTOLQNTFLDRSS 28
 1 DRVAIYEEFLMRTNGTOLQNTFLDRSS 28

RESULT 2
 ADJ66729
 ID ADJ66729 standard; protein; 108 AA.
 AC ADJ66729;
 XX
 XX 06-MAY-2004 (first entry)
 XX
 DE Human Muc16 epitope amino acid sequence.
 XX
 KW monoclonal antibody; epitope; non-shed extracellular portion;
 KW shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
 KW breast cancer; ovarian cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO2004005470-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 03-JUL-2003; 2003WO-US020907.
 XX
 PR 03-JUL-2002; 2002US-0393094P.
 XX
 PA (IMMU-) IMMUNOGEN INC.
 XX
 PI Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,
 PI Vater CA;
 DR WPI; 2004-091350/09.
 XX
 PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
 PT treating or monitoring malignancies, such as breast or ovarian cancer.
 XX
 PS Claim 13; SEQ ID NO 2; 113pp; English.
 XX
 CC This invention relates to a novel isolated monoclonal antibody that
 CC specifically binds to an epitope of a non-shed extracellular portion of a
 CC shed antigen or of human Muc1 or Muc16 protein. The invention may be
 CC useful for the development of compounds with a cytostatic or for gene
 CC therapy. The composition and methods are useful in diagnosing, treating
 CC or monitoring malignancies, such as breast or ovarian cancer. The present
 CC sequence is that of a muc epitope of the invention.

Sequence 108 AA;

Query Match 100.0%; Score 141; DB 8; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.6e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DRVAIYEEFLMRTNGTOLQNTFLDRSS 28
 60 DRVAIYEEFLMRTNGTOLQNTFLDRSS 87

RESULT 3
 ADJ66740
 ID ADJ66740 standard; protein; 108 AA.
 AC ADJ66740;
 XX
 XX 06-MAY-2004 (first entry)
 XX
 DE Human Muc16 GST fusion protein amino acid sequence SeqID13.

XX monoclonal antibody; epitope; non-shed extracellular portion;
 KW shed antigen; human; Muc1; Muc16; cytostatic; gene therapy; malignancy;
 KW breast cancer; ovarian cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO2004005470-A2.
 XX
 PD 15-JAN-2004.
 XX
 PF 03-JUL-2003; 2003WO-US020907.
 XX
 PR 03-JUL-2002; 2002US-0393094P.
 XX
 PA (IMMU-) IMMUNOGEN INC.
 XX
 PI Payne G, Chittenden T, Goldmakher V, Chun P, Sneider-Mulready K,
 PI Vater CA;
 DR WPI; 2004-091350/09.
 XX
 PT New antibody to non-shed human Muc1 or Muc16, useful for diagnosing,
 PT treating or monitoring malignancies, such as breast or ovarian cancer.
 XX
 PS Disclosure; SEQ ID NO 13; 113pp; English.
 XX
 CC This invention relates to a novel isolated monoclonal antibody that
 CC specifically binds to an epitope of a non-shed extracellular portion of a
 CC shed antigen or of human Muc1 or Muc16 protein. The invention may be
 CC useful for the development of compounds with a cytostatic or for gene
 CC therapy. The composition and methods are useful in diagnosing, treating
 CC or monitoring malignancies, such as breast or ovarian cancer. The present
 CC sequence is that of the human Muc16 GST fusion protein which was used in
 CC the exemplification of the invention.

Sequence 108 AA;

Query Match 100.0%; Score 141; DB 8; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.6e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 DRVAIYEEFLMRTNGTOLQNTFLDRSS 28
 60 DRVAIYEEFLMRTNGTOLQNTFLDRSS 87

RESULT 4
 ADM30756
 ID ADM30756 standard; protein; 174 AA.
 AC ADM30756;
 XX
 XX 20-MAY-2004 (first entry)
 XX
 DE Human cancer linked protein referred to as 192292 SeqID 4.
 XX
 KW human; cancer-related gene; screening assay; immuno-conjugate;
 KW cytostatic; immunostimulant; cancer; ovarian cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO2003075854-A2.
 XX
 PD 18-SEP-2003.
 XX
 PF 07-MAR-2003; 2003WO-US007147.
 XX
 PR 07-MAR-2002; 2002US-0362527P.
 XX
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Ebner R, Strovel JW;

XX WPI; 2003-748320/70.
DR N-PSDB; ADM30755.
XX
PT Identifying an agent that modulates the activity of any of three specific
PT cancer-related genes, potentially useful in treating (ovarian) cancer,
PT comprises detecting a difference in expression of the gene in the
PT presence of the agent.
XX
PS Claim 11; SEQ ID NO 4; 57bp; English.
XX
CC This invention relates to a novel method of identifying an agent that
CC modulates the activity of a cancer-related gene. Specifically, it refers
CC to a screening assay to identify potential antitumor agents, as well as
CC methods to assess the cancerous state of a cell. The present invention
CC describes antibodies against the expression products of these cancer-
CC related genes that are capable of targeting a cancerous cell in vivo.
CC Furthermore it provides immuno-conjugates containing such antibodies,
CC which can be used to deliver target therapeutics thereto. Accordingly,
CC these compositions that exhibit cytostatic and immunostimulant activities
CC are useful for preventing or treating cancer in humans and in particular
CC the treatment of ovarian cancer. This polypeptide sequence is a protein
CC encoded by a human cancer-linked gene of the invention.
SQ
SQ Sequence 174 AA;
XX
Query Match 100.0%; Score 141; DB 7; Length 174;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 DRVAIYEEFLRMTRNGTQLQNFLLDRSS 28
DB 70 DRVAIYEEFLRMTRNGTQLQNFLLDRSS 97
XX
RESULT 5
ABP30979
ID ABP30979 standard; protein; 178 AA.
XX
XX ABP30979;
XX
DT 02-JUL-2002 (first entry)
XX
DE Predicted extracellular domain of 0772P.
XX
KW Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
OS Homo sapiens.
XX
PN WO200206317-A2.
XX
PD 24-JAN-2002.
XX
PF 17-JUL-2001; 2001WO-US022635.
XX
PI 17-JUL-2000; 2000US-00617747.
XX
PR 10-AUG-2000; 2000US-00636801.
XX
PR 20-SEP-2000; 2000US-00667857.
XX
PR 04-APR-2001; 2001US-00827271.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvyck TS, Carter D, Hill P, Albone E;
XX
DR WPI; 2002-164781/21.
XX
PT Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
XX Example 13; Page 361-362; 408bp; English.

XX
CC This invention relates to polypeptides comprising an immunogenic portion
CC of an ovarian carcinoma protein which acts as an immunostimulant and is
CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations and antigen presenting cells that express
CC the polypeptides are useful for stimulating an immune response in a
CC patient and treating ovarian cancer. This sequence represents protein
CC related to the invention
XX
SQ
SQ Sequence 178 AA;
XX
Query Match 100.0%; Score 141; DB 5; Length 178;
Best Local Similarity 100.0%; Pred. No. 6.4e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 DRVAIYEEFLRMTRNGTQLQNFLLDRSS 28
DB 129 DRVAIYEEFLRMTRNGTQLQNFLLDRSS 156
XX
RESULT 6
ADA08642
ID ADA08642 standard; protein; 178 AA.
XX
XX ADA08642;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human 0772P partial protein #12.
XX
KW human; gene therapy; ovarian cancer; cancer.
XX
OS Homo sapiens.
XX
PN US2003091580-A1.
XX
PD 15-MAY-2003.
XX
PI 17-JUL-2001; 2001US-00907969.
XX
PR 18-JUN-2001; 2001US-00884441.
XX
PA (MITC/) MITCHAM J L.
PA (KING/) KING G E.
PA (ALGA/) ALGATE P A.
PA (FLIN/) FLING S P.
PA (RETT/) RETTER M W.
PA (FANG/) FANGER G R.
PA (REED/) REED S G.
PA (VEDV/) VEDDYCK T S.
PA (CART/) CARTER D.
PA (HILL/) HILL P.
PA (ALBO/) ALBONE E.
XX
PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Veddyck TS, Carter D, Hill P, Albone E;
XX
DR WPI; 2003-532352/50.
XX
PT New isolated 0772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
PT cancer.
XX
PS Disclosure; Page 28; 371bp; English.
XX
CC The invention relates to an isolated 0772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma cDNAs and protein cDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma

CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 141; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 6,4e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DRVAIYEEFLMTRNGTOLQNFLLDRSS 28
 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156
 Db 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156
 RESULT 7
 ADF08985
 ID ADF08985 standard; protein; 178 AA.
 XX ADF08985;
 AC
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Secreted ovarian carcinoma antigen seqid 489.
 XX
 KW gene therapy; protein therapy; vaccine; antibody inhibition;
 KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
 KW secreted ovarian carcinoma antigen.
 XX
 OS Homo sapiens.
 XX
 PN US2003124140-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 17-JUL-2002; 2002US-00198053.
 XX
 PR 17-DEC-1998; 98US-002155681.
 PR 17-DEC-1998; 98US-00216003.
 PR 23-JUN-1999; 99US-00338933.
 PR 24-SEP-1999; 99US-00404879.
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 PR 17-JUL-2001; 2001US-00907969.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Bangur CS, Retter MW, Fanger GR, Hill P;
 DR WPI; 2003-897152/82.
 XX
 PT Oncogenic nucleic acids useful for the prevention, diagnosis and
 PT treatment of breast cancer.
 XX
 PS Example 13; SEQ ID NO 489; 399pp; English.
 CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancer. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patient's own production of (II).
 CC Additionally, (II) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigens in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and as
 CC antagonists may be used to regulate expression and activity and as

CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunosassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.
 XX
 SQ Sequence 178 AA;
 Query Match 100.0%; Score 141; DB 7; Length 178;
 Best Local Similarity 100.0%; Pred. No. 6,4e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DRVAIYEEFLMTRNGTOLQNFLLDRSS 28
 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156
 Db 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156
 RESULT 8
 ABP30978
 ID ABP30978 standard; protein; 233 AA.
 XX ABP30978;
 AC
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Extracellular and transmembranal regions of 0772P.
 XX
 KW Human; immunostimulant; cyostatic; cancer; ovarian carcinoma.
 KW Human; immunostimulant; cyostatic; cancer; ovarian carcinoma.
 XX
 OS Homo sapiens.
 XX
 PN WO200206317-A2.
 XX
 PD 24-JAN-2002.
 XX
 PF 17-JUL-2001; 2001WO-US022635.
 XX
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00884441.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
 PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
 DR WPI; 2002-164781/21.
 XX
 PT Polypeptides comprising an immunogenic portion of an ovarian carcinoma
 PT protein or its variants, useful for stimulating an immune response in a
 PT patient and treating ovarian cancer.
 XX
 PS Example 13; Page 361; 408pp; English.
 CC This invention relates to polypeptides comprising an immunogenic portion
 CC of an ovarian carcinoma protein which acts as an immunostimulant and is
 CC cytostatic. The polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations and antigen presenting cells that express
 CC the polypeptides are useful for stimulating an immune response in a
 CC patient and treating ovarian cancer. This sequence represents protein
 CC related to the invention
 CC
 SQ Sequence 233 AA;
 Query Match 100.0%; Score 141; DB 5; Length 233;
 Best Local Similarity 100.0%; Pred. No. 8,8e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DRVAIYEEFLMTRNGTOLQNFLLDRSS 28
 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156
 Db 129 DRVAIYEEFLMTRNGTOLQNFLLDRSS 156

RESULT 9
 ID ADA08641 standard; protein; 233 AA.
 XX
 AC ADA08641;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human O772P partial protein #11.
 XX
 DE human; gene therapy; ovarian cancer; cancer.
 XX
 OS Homo sapiens.
 XX
 PN US2003091580-A1.
 XX
 PD 15-MAY-2003.
 XX
 PF 17-JUL-2001; 2001US-00907969.
 XX
 PR 18-JUN-2001; 2001US-0084441.
 XX
 PA (MITC/) MITCHAM J L.
 PA (KING/) KING G E.
 PA (ALGA/) ALGATE P A.
 PA (FLIN/) FLING S P.
 PA (RETT/) RETTER M W.
 PA (FANG/) FANGER G R.
 PA (REED/) REED S G.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (HILL/) HILL P.
 PA (ALBO/) ALBONE E.
 XX
 PI Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
 PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
 XX
 DR WPI; 2003-532352/50.
 XX
 PT New isolated O772P polypeptides and polynucleotides, useful in gene
 PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
 PT cancer.
 XX
 PS Example 13; Page 28; 371pp; English.
 XX
 CC The invention relates to an isolated O772P polypeptide, which has the
 CC structure fully defined in the specification. The composition containing
 CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
 CC or antigen presenting cells are useful for stimulating an immune response
 CC and treating ovarian cancer. Detecting the presence of the
 CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
 CC carcinoma cDNAs and protein cDNAs were identified using microarray
 CC technology. The present sequence represents a human ovarian carcinoma
 CC antigen.
 CC
 SQ Sequence 233 AA;
 XX
 Query Match 100.0%; Score 141; DB 7; Length 233;
 Best Local Similarity 100.0%; Pred. No. 8.8e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DRVAIYEEFLRMTRNGTQLONFLLDRSS 28
 Db 129 DRVAIYEEFLRMTRNGTQLONFLLDRSS 156
 XX
 RESULT 10
 ID ADF08984 standard; peptide; 233 AA.
 XX
 AC ADF08984;
 XX
 DT 12-FEB-2004 (first entry)

XX
 DE Secreted ovarian carcinoma antigen segid 488.
 XX
 KW gene therapy; protein therapy; vaccine; antibody inhibition;
 KW breast cancer; restorative therapy; diagnostic agent; immunoassay;
 KW secreted ovarian carcinoma antigen.
 XX
 OS Homo sapiens.
 XX
 PN US2003124140-A1.
 XX
 PD 03-JUL-2003.
 XX
 PF 17-JUL-2002; 2002US-00198053.
 XX
 PR 17-DEC-1998; 98US-00215681.
 PR 17-DEC-1998; 98US-00216003.
 PR 23-JUN-1999; 99US-00338933.
 PR 24-SEP-1999; 99US-00404879.
 PR 17-JUL-2000; 2000US-00617747.
 PR 10-AUG-2000; 2000US-00636801.
 PR 20-SEP-2000; 2000US-00667857.
 PR 04-APR-2001; 2001US-00827271.
 PR 18-JUN-2001; 2001US-00864441.
 PR 17-JUL-2001; 2001US-00907969.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Bangur CS, Retter MW, Fanger GR, Hill P;
 XX
 DR WPI; 2003-897152/82.
 XX
 PT Oncogenic nucleic acids useful for the prevention, diagnosis and
 PT treatment of breast cancer.
 XX
 PS Example 13; SEQ ID NO 488; 399pp; English.
 XX
 CC The invention describes nucleic acids (I) and the polypeptides (II) they
 CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
 CC treating diseases related to their aberrant expression i.e. breast
 CC cancers. For example, (I) and (II) may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of (II) by expressing
 CC inactive proteins or to supplement the patient's own production of (II).
 CC Additionally, (I) may be used to produce (II), by inserting (I) into a
 CC host cell and culturing the cell to express the protein (II). (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The host cell may also be used as antigens in the production of
 CC antibodies against (II) and in assays to identify modulators of (II)'s
 CC expression and activity. The anti-(II) antibodies, agonists and
 CC antagonists may be used to regulate expression and activity and as
 CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
 CC immunoassay). This sequence represents a secreted ovarian carcinoma
 CC antigen.
 CC
 SQ Sequence 233 AA;
 XX
 Query Match 100.0%; Score 141; DB 7; Length 233;
 Best Local Similarity 100.0%; Pred. No. 8.8e-14;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 DRVAIYEEFLRMTRNGTQLONFLLDRSS 28
 Db 129 DRVAIYEEFLRMTRNGTQLONFLLDRSS 156
 XX
 RESULT 11
 ID ABUS4859 standard; protein; 284 AA.
 XX
 AC ABUS4859;
 XX

```

XX 12-MAR-2003 (first entry)
DT
XX
XX Human CA125 carboxy terminal domain.
DE
XX
XX Human; CA125; protein repeat; chromosome 19q 13.2; amino terminal domain;
KW amino terminal extension; carboxy terminal domain; vaccine; cancer;
XX ovarian cancer; carcinoma.
XX
XX Homo sapiens.
OS
XX MO200283866-A2.
PN
XX
XX 24-OCT-2002.
PD
XX
XX 12-APR-2002; 2002WO-US011734.
PF
XX
XX 17-APR-2001; 2001US-0284175P.
PR 19-JUN-2001; 2001US-0299380P.
PR 27-SEP-2001; 2001US-00965738.
PR 21-DEC-2001; 2001US-0345180P.
XX
XX (UYAR-) UNITV ARKANSAS.
PA
XX O'brien T, Beard J, Underwood L;
PI
XX
XX WPI; 2003-093013/08.
DR
XX
XX New CA125 molecules, useful as a gold standard for detecting and
PT monitoring the presence of CA125 antigen which can be used for
PT diagnosing, monitoring or treating patients with cancer or for developing
PT vaccine against cancer.
XX
XX Claim 1; Fig 9b; 694bp; English.
XX
XX The invention relates to a CA125 protein comprising: (a) an extracellular
XX amino terminal domain; (b) an amino terminal extension; (c) a multiple
XX repeat domain; and (d) a carboxy terminal domain. The extracellular amino
XX terminal domain comprises 5 genomic exons, the amino terminal extension
XX comprises 4 genomic exons, each repeat unit comprises 5 genomic exons and
XX the carboxy terminal domain comprises a transmembrane anchor with a short
XX cytoplasmic domain, and further comprises 9 genomic exons. The gene for
XX CA125 is located on human chromosome 19q 13.2. Also included are isolated
XX CA125 repeat domains, nucleic acids (including variants, homologues and
XX degenerate versions) encoding CA125 proteins or repeat units, a vector
XX comprising the nucleic acid, a cultured cell comprising the vector, a
XX method of expressing CA125 antigen in a cell, the amino acid sequences of
XX the CA125 repeat units (or their variants, fragments or sequences 50%
XX identical to them), a purified antibody that selectively binds to an
XX epitope in the receptor-binding domain of CA125 protein, a diagnostic for
XX detecting and monitoring the presence of CA125 antigen (comprising
XX recombinant CA125 having at least one repeat unit of the CA125 repeat
XX domain including epitope binding sites), a therapeutic vaccine to treat
XX mammals with elevated CA125 antigen levels or at risk of developing a
XX disease or disease recurrence associated with elevated CA125 antigen
XX levels (comprising recombinant CA125 repeat domains including epitope
XX binding sites) and an antisense oligonucleotide that inhibits the
XX expression of CA125. The CA125 molecule, particularly the multiple repeat
XX domains are useful as a gold standard for detecting and monitoring the
XX presence of CA125 antigen, which can be used for diagnosing, monitoring
XX or treating patients with ovarian cancer and other carcinomas where CA125
XX is expressed. The molecules are also useful for developing a vaccine
XX against cancer. The present sequence is a CA125 repeat protein
SQ
Sequence 284 AA;
Query Match 100.0%; Score 141; DB 6; Length 284;
Best Local Similarity 100.0%; Pred. No. 1,1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DRVAIYEEFLMTRNGTQLQNFLLDRSS 28
DB 180 DRVAIYEEFLMTRNGTQLQNFLLDRSS 207

```

```

RESULT 12
ABP31025
ID ABP31025 standard; protein; 318 AA.
XX
XX AC ABP31025;
XX
XX 02-JUL-2002 (first entry)
DT
XX
XX Amino acid sequence of the 3' constant region of 0772P.
DE
XX
XX Human; immunostimulant; cytostatic; cancer; ovarian carcinoma.
XX
XX Homo sapiens.
OS
XX MO200206317-A2.
PN
XX
XX 24-JAN-2002.
PD
XX
XX 17-JUL-2001; 2001WO-US022635.
PF
XX
XX 17-JUL-2000; 2000US-00617747.
PR 10-AUG-2000; 2000US-00636801.
PR 20-SEP-2000; 2000US-00667857.
PR 04-APR-2001; 2001US-00827271.
PR 18-JUN-2001; 2001US-00884441.
XX
XX (CORI-) CORIXA CORP.
PA
XX Mitcham UL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2002-164781/21.
DR
XX
XX Polypeptides comprising an immunogenic portion of an ovarian carcinoma
PT protein or its variants, useful for stimulating an immune response in a
PT patient and treating ovarian cancer.
XX
XX Claim 3; Page 399-400; 408bp; English.
XX
XX This invention relates to polypeptides comprising an immunogenic portion
XX of an ovarian carcinoma protein which acts as an immunostimulant and is
XX cytostatic. The polypeptides, polynucleotides, antibodies, fusion
XX proteins, T cell populations and antigen presenting cells that express
XX the polypeptides are useful for stimulating an immune response in a
XX patient and treating ovarian cancer. This sequence represents protein
XX related to the invention
SQ
Sequence 318 AA;
Query Match 100.0%; Score 141; DB 5; Length 318;
Best Local Similarity 100.0%; Pred. No. 1,3e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DRVAIYEEFLMTRNGTQLQNFLLDRSS 28
DB 214 DRVAIYEEFLMTRNGTQLQNFLLDRSS 241

```

```

RESULT 13
ADA08747
ID ADA08747 standard; protein; 318 AA.
XX
XX AC ADA08747;
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Human O772P constant region consensus sequence #1.
DE
XX
XX human; gene therapy; ovarian cancer; cancer.
XX
XX Homo sapiens.
OS

```

```
XX US2003091580-A1.
PN
XX
XX 15-MAY-2003.
PD
XX
XX 17-JUL-2001; 2001US-00907969.
PR
XX
XX 18-JUN-2001; 2001US-00884441.
PR
XX
XX (MITC/) MITCHAM J L.
PA (KING/) KING G B.
PA (ALGA/) ALGATE P A.
PA (FLIN/) FLING S P.
PA (RETT/) RETTER M W.
PA (FANG/) FANGER G R.
PA (REED/) REED S G.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (HILL/) HILL P.
PA (ALBO/) ALBONE E.
XX
XX Mitcham JL, King GE, Algate PA, Fling SP, Retter MW, Fanger GR;
PI Reed SG, Vedvick TS, Carter D, Hill P, Albone E;
XX
XX WPI; 2003-532352/50.
XX
XX New isolated 0772P polypeptides and polynucleotides, useful in gene
PT therapy, particularly for treating or diagnosing cancer, e.g. ovarian
XX cancer.
XX
XX Claim 3; SEQ ID NO 594; 371pp; English.
XX
XX The invention relates to an isolated 0772P polypeptide, which has the
CC structure fully defined in the specification. The composition containing
CC the polypeptides, polynucleotides, fusion proteins, antibodies, T-cells
CC or antigen presenting cells are useful for stimulating an immune response
CC and treating ovarian cancer. Detecting the presence of the
CC polynucleotides and polypeptides is useful for diagnosing cancer. Ovarian
CC carcinoma CDNAs and protein CDNAs were identified using microarray
CC technology. The present sequence represents a human ovarian carcinoma
CC antigen.
XX
XX
XX Sequence 318 AA;
SQ
XX
XX Query Match 100.0%; Score 141; DB 7; Length 318;
XX Best Local Similarity 100.0%; Pred.No.1.3e-13;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
QY |||||
Db 214 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 241
XX
XX
XX RESULT 14
XX ADF09090
XX ID ADF09090 standard; protein; 318 AA.
XX
XX ADF09090;
XX
XX 12-FEB-2004 (first entry)
XX
XX Secreted ovarian carcinoma antigen seqid 594.
XX
XX gene therapy; protein therapy; vaccine; antibody inhibition;
XX breast cancer; restorative therapy; diagnostic agent; immunoassay;
XX secreted ovarian carcinoma antigen.
XX
XX Homo sapiens.
XX
XX US2003124140-A1.
XX
XX 03-JUL-2003.
XX
XX
```

```
PF 17-JUL-2002; 2002US-00198053.
XX
XX 17-DEC-1998; 98US-00215681.
PR 17-DEC-1998; 98US-00216003.
PR 23-JUN-1999; 99US-00338933.
PR 24-SEP-1999; 99US-00404879.
PR 17-JUL-2000; 2000US-00617747.
PR 10-AUG-2000; 2000US-00636801.
PR 20-SEP-2000; 2000US-00667857.
PR 04-APR-2001; 2001US-00827271.
PR 18-JUN-2001; 2001US-00884441.
PR 17-JUL-2001; 2001US-00907969.
XX
XX (CORI-) CORIXA CORP.
XX
XX
XX Bangur CS, Retter MW, Fanger GR, Hill P;
PI WPI; 2003-897152/82.
XX
XX N-PSDB; ADF09064.
XX
XX Oncogenic nucleic acids useful for the prevention, diagnosis and
PT treatment of breast cancer.
XX
XX Example 16; SEQ ID NO 594; 399pp; English.
XX
XX
XX The invention describes nucleic acids (I) and the polypeptides (II) they
CC encode. The nucleic acids (I) may be used for preventing, diagnosing and
CC treating diseases related to their aberrant expression i.e. breast
CC cancers. For example, (I) and (II) may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of (II) by expressing
CC inactive proteins or to supplement the patients own production of (II).
CC Additionally, (I) may be used to produce (II), by inserting (I) into a
CC host cell and culturing the cell to express the protein (II). (I) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. The host cell may also be used as antigens in the production of
CC antibodies against (II) and in assays to identify modulators of (II)'s
CC expression and activity. The anti-(II) antibodies, agonists and
CC antagonists may be used to regulate expression and activity and as
CC diagnostic agents for detecting the presence of (II) in samples (e.g. by
CC immunoassay). This sequence represents a secreted ovarian carcinoma
CC antigen.
XX
XX
XX Sequence 318 AA;
SQ
XX
XX Query Match 100.0%; Score 141; DB 7; Length 318;
XX Best Local Similarity 100.0%; Pred.No.1.3e-13;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
QY |||||
Db 214 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 241
XX
XX
XX RESULT 15
XX AAE12631
XX ID AAE12631 standard; protein; 367 AA.
XX
XX AAE12631;
XX
XX 03-JAN-2002 (first entry)
XX
XX Human gene 1 encoded secreted protein fragment, SEQ ID NO: 27.
XX
XX Human; ovarian cancer antigen; proliferative disorder; cancer; tumour;
XX respiratory system disorder; asthma; haematopoietic disorder; skin aging;
XX immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
XX rheumatoid arthritis; inflammation; neurological disorder; septic shock;
XX Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
XX atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
XX epithelial cell proliferation; transplantation; chemotaxis; infection;
```

KM food additive; wound healing; endocrine disorder; kidney disorder;
 KW gene therapy; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN MO200170804-A1.
 XX
 PD 27-SEP-2001.
 XX
 PF 16-MAR-2001; 2001WO-US008585.
 XX
 PR 17-MAR-2000; 2000US-0190076P.
 PR 23-AUG-2000; 2000US-0227009P.
 XX
 PA (HUMA-) HUMAN GENOME SCT INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2001-639119/73.
 XX
 PT Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
 PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
 PT metastases.
 XX
 PS Disclosure; Page 9; 427pp; English.
 XX
 CC AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
 CC ovarian cancer associated protein (collectively known as ovarian cancer
 CC antigens) genes; and AAE12623-AAE12629 represent the proteins they
 CC encode. AAE12630-AAE12638 represent human ovarian cancer antigenic
 CC fragments or variants. Ovarian cancer antigens and their corresponding
 CC DNAs are used in the prevention, diagnosis and treatment of diseases
 CC associated with their inappropriate expression. These disorders include
 CC proliferative disorders, cancer, tumours, respiratory system disorders,
 CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
 CC skin disorders, autoimmune diseases (e.g., rheumatoid arthritis),
 CC inflammation, allergies, neurological disorders (e.g., Alzheimer's
 CC disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, endocrine disorders and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, to identify their cognate ligands or binding
 CC partners, in chemotaxis and can be used as a food additive. Antibodies
 CC specific for a protein of the invention can be used in alleviating
 CC symptoms associated with the disorders mentioned above and in diagnostic
 CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
 CC the invention is used in gene therapy. The present sequence represents a
 CC human ovarian cancer antigenic fragment of the invention
 XX
 SQ Sequence 367 AA;
 XX
 Query Match 100.0%; Score 141; DB 4; Length 367;
 Best Local Similarity 100.0%; Pred. No. 1.5e-13;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 DRVAIYEEFLRMTRNGTOLQNFPTLDRSS 28
 |||||
 DB 334 DRVAIYEEFLRMTRNGTOLQNFPTLDRSS 361

Search completed: October 27, 2005, 18:44:25
 Job time : 96.893 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:03:42 ; Search time 24.3535 Seconds
(without alignments)
85.826 Million cell updates/sec

Title: US-10-612-090-18

Perfect score: 141
Sequence: 1 DRVAIYEEFLRMTRNGTQLONFLLDRSS 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	100.0	438	4	US-09-404-879A-390 Sequence 390, App
2	141	100.0	438	4	US-09-667-857-390 Sequence 390, App
3	141	100.0	833	4	US-09-404-879A-389 Sequence 389, App
4	141	100.0	833	4	US-09-667-857-389 Sequence 389, App
5	141	100.0	914	4	US-09-404-879A-312 Sequence 312, App
6	141	100.0	914	4	US-09-338-933-312 Sequence 312, App
7	141	100.0	914	4	US-09-667-857-312 Sequence 312, App
8	56	39.7	772	4	US-09-404-879A-388 Sequence 388, App
9	56	39.7	772	4	US-09-667-857-388 Sequence 388, App
10	48.5	34.4	263	4	US-09-710-279-332 Sequence 332, App
11	48.5	34.4	274	3	US-09-134-001C-3921 Sequence 3921, App
12	47	33.3	514	3	US-09-066-047-3 Sequence 3, App1
13	47	33.3	2088	4	US-09-802-540-12906 Sequence 12906, A
14	46.5	33.0	693	4	US-09-818-780-68 Sequence 68, App1
15	46	32.6	115	4	US-09-107-532A-5481 Sequence 5481, App
16	45.5	32.3	433	4	US-09-328-352-7948 Sequence 7948, App
17	45	31.9	245	4	US-09-270-767-45297 Sequence 45297, A
18	45	31.9	255	4	US-09-270-767-32769 Sequence 32769, A
19	45	31.9	255	4	US-09-370-767-47986 Sequence 47986, A
20	45	31.9	275	4	US-09-338-092-348 Sequence 348, App
21	44.5	31.6	145	3	US-08-946-329A-46 Sequence 46, App1
22	44.5	31.6	145	3	US-08-946-329A-69 Sequence 69, App1
23	44.5	31.6	297	3	US-09-058-489-10 Sequence 10, App1
24	44.5	31.6	376	4	US-09-849-016-8106 Sequence 8106, App
25	44.5	31.6	404	4	US-09-949-016-6202 Sequence 6202, App
26	44.5	31.6	416	4	US-09-949-016-6201 Sequence 6201, App
27	44.5	31.6	433	4	US-09-949-016-10900 Sequence 10900, A

28	44.5	31.6	471	1	US-08-176-427B-11 Sequence 11, App1
29	44.5	31.6	471	2	US-08-356-060A-34 Sequence 34, App1
30	44.5	31.6	471	3	US-08-460-900C-34 Sequence 34, App1
31	44.5	31.6	471	3	US-08-757-230A-8 Sequence 8, App1
32	44.5	31.6	471	3	US-08-674-509B-34 Sequence 34, App1
33	44.5	31.6	471	3	US-08-954-698-34 Sequence 34, App1
34	44.5	31.6	471	3	US-08-957-874-34 Sequence 34, App1
35	44.5	31.6	471	4	US-09-325-256-26 Sequence 26, App1
36	44.5	31.6	471	4	US-08-700-393-8 Sequence 8, App1
37	44.5	31.6	471	4	US-09-639-693-34 Sequence 34, App1
38	44.5	31.6	471	4	US-09-448-188-34 Sequence 34, App1
39	44.5	31.6	471	4	US-08-954-128-34 Sequence 34, App1
40	44.5	31.6	471	4	US-09-704-917-20 Sequence 20, App1
41	44.5	31.6	471	4	US-08-954-740-34 Sequence 34, App1
42	44.5	31.6	471	4	US-09-151-999-20 Sequence 20, App1
43	44.5	31.6	471	4	US-09-736-476-34 Sequence 34, App1
44	44.5	31.6	471	4	US-09-418-221-18 Sequence 18, App1
45	44.5	31.6	560	2	US-08-948-569A-10 Sequence 10, App1

ALIGNMENTS

```
RESULT 1
; US-09-404-879A-390
; Sequence 390, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-404-879A-390

Query Match      100.0%; Score 141; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1
|||||
Db      334 DRVAIYEEFLRMTRNGTQLONFLLDRSS 361

RESULT 2
; US-09-667-857-390
; Sequence 390, Application US/09667857
; Patent No. 6689654
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedicik, Thomas S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
```

```
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 390
/ LENGTH: 438
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-667-857-390

Query Match          100.0%; Score 141; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 334 DRAVYEEFLMTRNGTOLQNFLLDRSS 361

RESULT 3
US-09-404-879A-389
/ Sequence 389, Application US/09404879A
/ Patent No. 6468546
/ GENERAL INFORMATION:
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: King, Gordon E.
/ APPLICANT: Algate, Paul A.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.462C2
/ CURRENT APPLICATION NUMBER: US/09/404,879A
/ CURRENT FILING DATE: 1999-09-24
/ NUMBER OF SEQ ID NOS: 393
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 389
/ LENGTH: 833
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-404-879A-389

Query Match          100.0%; Score 141; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 729 DRAVYEEFLMTRNGTOLQNFLLDRSS 756

RESULT 4
US-09-667-857-389
/ Sequence 389, Application US/09667857
/ Patent No. 6699664
/ GENERAL INFORMATION:
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: King, Gordon E.
/ APPLICANT: Algate, Paul A.
/ APPLICANT: Flings, Steven P.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Fanger, Gary Richard
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Vedvick, Thomas S.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.462C5
/ CURRENT APPLICATION NUMBER: US/09/667,857
/ CURRENT FILING DATE: 2000-09-20
/ NUMBER OF SEQ ID NOS: 455
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 389
/ LENGTH: 833
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-667-857-389
```

```
Query Match          100.0%; Score 141; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 2.3e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 729 DRAVYEEFLMTRNGTOLQNFLLDRSS 756

RESULT 5
US-09-404-879A-312
/ Sequence 312, Application US/09404879A
/ Patent No. 6468546
/ GENERAL INFORMATION:
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: King, Gordon E.
/ APPLICANT: Algate, Paul A.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.462C2
/ CURRENT APPLICATION NUMBER: US/09/404,879A
/ CURRENT FILING DATE: 1999-09-24
/ NUMBER OF SEQ ID NOS: 393
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 312
/ LENGTH: 914
/ TYPE: PRT
/ ORGANISM: Homo sapien
US-09-404-879A-312

Query Match          100.0%; Score 141; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 810 DRAVYEEFLMTRNGTOLQNFLLDRSS 837

RESULT 6
US-09-338-933-312
/ Sequence 312, Application US/09338933
/ Patent No. 648931
/ GENERAL INFORMATION:
/ APPLICANT: Mitcham, Jennifer Lynn
/ APPLICANT: King, Gordon E.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
/ TITLE OF INVENTION: OVARIAN CANCER
/ FILE REFERENCE: 210121.462C1
/ CURRENT APPLICATION NUMBER: US/09/338,933
/ CURRENT FILING DATE: 1999-06-23
/ NUMBER OF SEQ ID NOS: 312
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 312
/ LENGTH: 914
/ TYPE: PRT
/ ORGANISM: Homo sapien
US-09-338-933-312

Query Match          100.0%; Score 141; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 810 DRAVYEEFLMTRNGTOLQNFLLDRSS 837

RESULT 7
US-09-667-857-312
/ Sequence 312, Application US/09667857
/ Patent No. 6699664
/ GENERAL INFORMATION:
```

```

; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Ranger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 312
; LENGTH: 914
; TYPE: PRF
; ORGANISM: Homo sapien
US-09-667-857-312
```

```

Query Match          100.0%; Score 141; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 2,6e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 28
      |||||:|||||:|||||:|||||:
DB 810 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 837
```

```

RESULT 8
US-09-404-879A-388
; Sequence 388, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 388
; LENGTH: 772
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-404-879A-388
```

```

Query Match          39.7%; Score 56; DB 4; Length 772;
Best Local Similarity 42.9%; Pred. No. 1.6;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 28
      |||||:|||||:|||||:|||||:
DB 281 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 308
```

```

RESULT 9
US-09-667-857-388
; Sequence 388, Application US/09667857
; Patent No. 6699664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Ranger, Gary Richard
; APPLICANT: Reed, Steven G.
```

```

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 388
; LENGTH: 772
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-667-857-388
```

```

Query Match          39.7%; Score 56; DB 4; Length 772;
Best Local Similarity 42.9%; Pred. No. 1.6;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 28
      |||||:|||||:|||||:|||||:
DB 281 DRVAIYEEFLRMTRNGTOLQNFITLDRSS 308
```

```

RESULT 10
US-09-710-279-332
; Sequence 332, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 332
; LENGTH: 263
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-332
```

```

Query Match          34.4%; Score 48.5; DB 4; Length 263;
Best Local Similarity 37.9%; Pred. No. 7.1;
Matches 11; Conservative 7; Mismatches 8; Indels 3; Gaps 1;
```

```
QY 3 VAIIYEEFLRMTRNGTOLQNFITLDRSS 28
      |||||:|||||:|||||:|||||:
DB 33 VAIIEFKSPYKLTTRDGEKTLQYALQRES 61
```

```

RESULT 11
US-09-134-001C-3921
; Sequence 3921, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3921
```

```

; LENGTH: 274
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3921

```

Query Match	34.4%	Score 48.5;	DB 3;	Length 274;
Best Local Similarity	37.9%	Pred. No. 7.5;		
Matches 11; Conservative	7;	Mismatches 8;	Indels 3;	Gaps 1;

OY 3 VAIVE---EFLRMTRNGTOLONFTLDRSS 28
|::||::|:|:
Db 44 VKIEFSPYLKLTREDGEKLQVALQRES 72

RESULT 12
US-09-066-047-3
; Sequence 3, Application US/09066047A

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-066-047-3

```
Oy      3 VAIEEFLRMTRNGTOLQNF 24
          ||:|:|:|:|:|:|
Db      380 VAMLEQLIKYRDGIDLENN 401
```

RESULT 13
US-09-902-540-12906
; Sequence 12906, Application US/09902540
; Patent No. 6833447

Query Match	33.3%	Score 47	DB 4	Length 2088
Best Local Similarity	50.0%	Pred. No. 1.5e+02		
Matches 10	Conservative 1	Mismatches 9	Indels 0	Gaps 0

```

QY      5 IYEEFLRMTRNGTOLQNFTL 24
      :||| ||| |||
Db     299 VYETVLRHTRGRALDRFPL 318

```

RESULT 14
 US-09-818-780-68
 , Sequence 68, Application US/09818780
 , Patent No. 6677146
 , GENERAL INFORMATION:
 , APPLICANT: McHenry, Charles
 , TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III HOLOENZYME
 , FILE REFERENCE: 1794.0030004
 , CURRENT APPLICATION NUMBER: US/09/818,780
 , CURRENT FILING DATE: 2001-03-28
 , PRIOR APPLICATION NUMBER: US 60/192,736
 , PRIOR FILING DATE: 2000-03-28
 , NUMBER OF SEQ ID NOS: 98
 , SOFTWARE: PatentIn version 3.0
 , SEQ ID NO: 68
 , LENGTH: 693
 , TYPE: prt
 , ORGANISM: Thermus thermophilus
 , US-09-818-780-68

Query Match	33.0%	Score	46.5	DB	4	Length	693
Best Local	46.2%	Pred. NC	49				
Matches	12	Conservative	3	Mismatches	10	Indels	1
						Gaps	1

Qy 1 DRAVAYBEFLRMTNGTOLQNFYTLDR 26
||: |||: |||
Db 497 DRLNVEEELLRRAKEADLQDF-LDR 521

RESULT 15
US-09-107-532A-5481
; Sequence 5481, Application US/09107532A


```

STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Artifiello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5481:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...115
SEQUENCE DESCRIPTION: SEQ ID NO: 5481:
US-09-107-532A-5481

```

QY	3	V	A	I	E	E	F	L	K	M	T	R	N	G	T	O	L	N	F	T	L	D	R	S	28
				:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:		
Db	28	V	A	V	Y	Q	A	F	L	K	P	D	Y	G	T	K	L	I	S	T	N	G	T	53	

Query Match 33.6%; Score 46; DB 4;

Best Local Similarity 34.6%; Pred. No. 6.7;

Matches 9; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Length 115;

Search completed: October 27, 2005, 17:34:35
Job time : 25.3535 secs

(copy) name code SQL

Result No.	Score	Query Match	Length	DB	ID	Description
1	141	100.0	28	15	US-10-612-090-18	Sequence 18, Appl
2	141	100.0	108	15	US-10-612-090-2	Sequence 2, Appl
3	141	100.0	108	15	US-10-612-090-13	Sequence 13, Appl
4	141	100.0	174	9	US-10-383-368-4	Sequence 4, Appl
5	141	100.0	178	9	US-09-884-441-489	Sequence 49, App
6	141	100.0	178	10	US-09-907-969-689	Sequence 489, App
7	141	100.0	178	14	US-10-139-053-489	Sequence 489, App
8	141	100.0	178	17	US-10-860-790-889	Sequence 489, App
9	141	100.0	233	9	US-09-984-441-488	Sequence 488, App
10	141	100.0	233	10	US-09-907-969-488	Sequence 488, App
11	141	100.0	233	14	US-10-139-053-488	Sequence 488, App

12	141	100.0	233	17	US-10-860-790-488	Sequence 488, App
13	141	100.0	284	10	US-09-965-738-300	Sequence 300, App
14	141	100.0	318	10	US-09-907-969-594	Sequence 594, App
14	141	100.0	318	14	US-10-198-053-594	Sequence 594, App
15	141	100.0	318	17	US-10-860-790-594	Sequence 594, App
16	141	100.0	367	15	US-10-333-900-27	Sequence 27, App1
17	141	100.0	438	9	US-09-884-441-590	Sequence 390, App
18	141	100.0	438	9	US-09-884-441-483	Sequence 483, App
19	141	100.0	438	10	US-09-907-969-390	Sequence 390, App
20	141	100.0	438	10	US-09-907-969-483	Sequence 483, App
21	141	100.0	438	10	US-09-827-271-390	Sequence 390, App
22	141	100.0	438	14	US-10-198-053-390	Sequence 390, App
23	141	100.0	438	14	US-10-198-053-483	Sequence 483, App
24	141	100.0	438	15	US-10-333-900-19	Sequence 19, App1
25	141	100.0	438	17	US-10-860-790-483	Sequence 390, App
26	141	100.0	438	17	US-10-860-790-390	Sequence 483, App
27	141	100.0	438	17	US-10-860-790-483	Sequence 483, App
28	141	100.0	439	10	US-09-965-738-148	Sequence 148, App
29	141	100.0	526	15	US-10-333-900-30	Sequence 30, App1
30	141	100.0	545	14	US-10-343-2433-4	Sequence 4, App1i
31	141	100.0	583	14	US-10-142-515-4	Sequence 4, App1i
32	141	100.0	748	17	US-10-687-035-1	Sequence 2, App1i
33	141	100.0	809	17	US-10-687-035-2	Sequence 2, App1i
34	141	100.0	833	9	US-09-884-441-889	Sequence 389, App
35	141	100.0	833	10	US-09-907-969-389	Sequence 389, App
36	141	100.0	833	10	US-09-827-271-389	Sequence 389, App
37	141	100.0	833	14	US-10-198-053-389	Sequence 389, App
38	141	100.0	833	17	US-10-860-790-389	Sequence 389, App
39	141	100.0	914	9	US-09-778-320-206	Sequence 206, App
40	141	100.0	914	9	US-09-910-689-206	Sequence 206, App
41	141	100.0	914	9	US-09-884-441-512	Sequence 312, App
42	141	100.0	914	9	US-09-884-441-512	Sequence 478, App
43	141	100.0	914	10	US-09-907-969-312	Sequence 312, App
44	141	100.0	914	10	US-09-907-969-478	Sequence 478, App
45	141	100.0	914	10	US-09-827-271-312	Sequence 312, App

ALIGNMENTS

RESULT 1
 US-10-612-090-18
 ; Sequence 18, Application US/10612090
 ; Publication NO. US20040057952A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunogen, Inc.
 ; TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
 ; FILE REFERENCE: A8340
 ; CURRENT APPLICATION NUMBER: US/10/612,090
 ; CURRENT FILING DATE: 2003-07-03
 ; PRIOR APPLICATION NUMBER: US 60/393,094
 ; PRIOR FILING DATE: 2002-07-03
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 18
 ; LENGTH: 28
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-612-090-18

```

Query Match      100.0%; Score 141; DB 15; Length 28;
Best Local Similarity 100.0%; Pred. No. 9,2e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY      1 DRVAIYEEFLMTNNGTOLQNFITLDRSS 28
        |||||
Db       1 DRVAIYEEFLMTNNGTOLQNFITLDRSS 28
        |||||

RESULT 2
US-10-612-090-2
; Sequence 2, Application US/10612090
; Publication No. US20040057952x1
; GENERAL INFORMATION:

```

```

/ APPLICANT: Immunogen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: SeqIdn version 3.2
/ SEQ ID NO 2
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-612-090-2

Query Match      100.0%; Score 141; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 4,4e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DRVAIYEEFLMTRNGTOLQNFLLDRSS 28
      |||||||
Db      60 DRVAIYEEFLMTRNGTOLQNFLLDRSS 87

RESULT 3
US-10-612-090-13
/ Sequence 13, Application US/10612090
/ Publication No. US20040057952A1
/ GENERAL INFORMATION:
/ APPLICANT: Immunogen, Inc.
/ TITLE OF INVENTION: ANTIBODIES TO NON-SHED MUC1 AND MUC16, AND USES THEREOF
/ FILE REFERENCE: A8340
/ CURRENT APPLICATION NUMBER: US/10/612,090
/ CURRENT FILING DATE: 2003-07-03
/ PRIOR APPLICATION NUMBER: US 60/393,094
/ PRIOR FILING DATE: 2002-07-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 13
/ LENGTH: 108
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Fusion protein
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (1)..(1)
/ OTHER INFORMATION: Glutathione S-transferase fusion site
/ US-10-612-090-13

Query Match      100.0%; Score 141; DB 15; Length 108;
Best Local Similarity 100.0%; Pred. No. 4,4e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DRVAIYEEFLMTRNGTOLQNFLLDRSS 28
      |||||||
Db      60 DRVAIYEEFLMTRNGTOLQNFLLDRSS 87

RESULT 4
US-10-383-368-4
/ Sequence 4, Application US/10383368
/ Publication No. US20040002091A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul E.
/ APPLICANT: Ebner, Reinhard
/ TITLE OF INVENTION: Cancer-Linked Gene as Target for Chemotherapy
/ FILE REFERENCE: 689290-123
/ CURRENT APPLICATION NUMBER: US/10/383,368
/ CURRENT FILING DATE: 2003-03-07
/ PRIOR APPLICATION NUMBER: US 60/362,527
/ PRIOR FILING DATE: 2002-03-07

```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-383-368-4

Query Match          100.0%; Score 141; DB 15; Length 174;
Best Local Similarity 100.0%; Pred. No. 7,7e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  DRVAIYEEFLMRTNGTOLQNFILDRSS 28
      |||||||
DB      70  DRVAIYEEFLMRTNGTOLQNFILDRSS 97

RESULT 5
US-09-884-441-489
; Sequence 489, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Derrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-489

Query Match          100.0%; Score 141; DB 9; Length 178;
Best Local Similarity 100.0%; Pred. No. 8e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  DRVAIYEEFLMRTNGTOLQNFILDRSS 28
      |||||||
DB      129  DRVAIYEEFLMRTNGTOLQNFILDRSS 156

RESULT 6
US-09-907-969-489
; Sequence 489, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Hill, Paul
; APPLICANT: Albane, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-907-969-489

Query Match 100.0%; Score 141; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 8e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
Db 129 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 156

RESULT 7

US-10-198-053-489
; Sequence 489, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-489

Query Match 100.0%; Score 141; DB 14; Length 178;
Best Local Similarity 100.0%; Pred. No. 8e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
Db 129 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 156

RESULT 8

US-10-860-790-489
; Sequence 489, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-489

Query Match 100.0%; Score 141; DB 17; Length 178;
Best Local Similarity 100.0%; Pred. No. 8e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
Db 129 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 156

RESULT 9

US-09-884-441-488
; Sequence 488, Application US/09884441
; Patent No. US20020119158A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C7
; CURRENT APPLICATION NUMBER: US/09/884,441
; CURRENT FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-884-441-488

Query Match 100.0%; Score 141; DB 9; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
Db 129 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 156

RESULT 10

US-09-907-969-488
; Sequence 488, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-969-488

Query Match 100.0%; Score 141; DB 10; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
Db 129 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 156

RESULT 11

US-10-198-053-488
; Sequence 488, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-053-488

Query Match          100.0%; Score 141; DB 14; Length 233;
Best Local Similarity 100.0%; Pred. No. 1,1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 DRAVAYEEFLMTRNGTQLQNFLLDRSS 28
|||||
Db 129 DRAVAYEEFLMTRNGTQLQNFLLDRSS 156

RESULT 12
US-10-860-790-488
; Sequence 488, Application US/10860790
; Publication No. US20050031634A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C11
; CURRENT APPLICATION NUMBER: US/10/860,790
; CURRENT FILING DATE: 2004-06-02
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-860-790-488

Query Match          100.0%; Score 141; DB 17; Length 233;
Best Local Similarity 100.0%; Pred. No. 1,1e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 DRAVAYEEFLMTRNGTQLQNFLLDRSS 28
|||||
Db 129 DRAVAYEEFLMTRNGTQLQNFLLDRSS 156

RESULT 13
US-09-965-738-300
; Sequence 300, Application US/09965738
; Publication No. US2003014367A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy
; TITLE OF INVENTION: Repeat Sequences of the CA125 Gene and Their Use for Diagnostic A
; TITLE OF INVENTION: Therapeutic Interventions
; FILE REFERENCE: 40715-258841
; CURRENT APPLICATION NUMBER: US/09/965,738
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/284,175
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 300
; LENGTH: 284

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-738-300

Query Match          100.0%; Score 141; DB 10; Length 284;
Best Local Similarity 100.0%; Pred. No. 1,4e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 DRAVAYEEFLMTRNGTQLQNFLLDRSS 28
|||||
Db 180 DRAVAYEEFLMTRNGTQLQNFLLDRSS 207

RESULT 14
US-09-907-969-594
; Sequence 594, Application US/09907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedicik, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; CURRENT FILING DATE: 2001-07-17
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-09-907-969-594

Query Match          100.0%; Score 141; DB 10; Length 318;
Best Local Similarity 100.0%; Pred. No. 1,6e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 DRAVAYEEFLMTRNGTQLQNFLLDRSS 28
|||||
Db 214 DRAVAYEEFLMTRNGTQLQNFLLDRSS 241

RESULT 15
US-10-198-053-594
; Sequence 594, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 594
; LENGTH: 318

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: 136,248,268
; OTHER INFORMATION: Xaa = Any amino acid
US-10-198-053-594

```

```

Query Match          100.0%; Score 141; DB 14; Length 318;
Best Local Similarity . 100.0%; Pred. No. 1,6e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 28
          |||||
Db      214 DRVAIYEEFLRMTRNGTOLQNFLLDRSS 241

```

Search completed: October 27, 2005, 18:31:47
 Job time : 89.6 secs

The Negro Churchman

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 18:13:49 ; Search time 17.3209 Seconds
(without alignments)
155.538 Million cell updates/sec

Title: US-10-612-090-18

Perfect score: 141

Sequence: 1 DRVAIYEEFLRMTRNGTQLQNFPLDRSS 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_79:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.5	37.9	311	2 AG1271	hypothetical prote
2	53.5	37.9	311	2 A11633	hypothetical prote
3	49	34.8	540	1 T01989	calium-dependent
4	48	34.0	281	2 AD2312	hypothetical prote
5	48	34.0	384	2 G82670	general secretory
6	48	34.0	576	2 A84902	auxin-regulated pr
7	47.5	33.7	466	1 B43332	glutamate decarbox
8	47.5	33.7	466	1 S24234	glutamate decarbox
9	47.5	33.7	466	1 B91178	glutamate decarbox
10	47.5	33.7	466	2 F86024	glutamate decarbox
11	47.5	33.7	466	2 G85726	glutamate decarbox
12	47.5	33.7	466	2 B90891	glutamate decarbox
13	47	33.3	231	2 T32146	hypothetical prote
14	47	33.3	316	2 AB2868	hypothetical prote
15	47	33.3	316	2 F97644	UDP-hexose transfe
16	47	33.3	533	2 A86433	UDP-hexose transfe
17	47	33.3	2890	2 B71846	T518.19 protein -
18	47	33.3	2890	2 F64669	dna-directed RNA p
19	46.5	33.0	397	2 D96580	DNA-directed RNA p
20	46	32.6	113	2 H86296	hypothetical prote
21	46	32.6	249	2 T35589	F309.1 protein - A
22	46	32.6	514	2 T10938	probable secreted
23	46	32.6	725	1 Z1B222	calium-dependent
24	45.5	32.3	862	1 E84507	gene 1 protein - p
25	45	31.9	71	2 G97985	hypothetical prote
26	45	31.9	272	2 T15351	hypothetical prote
27	45	31.9	275	2 S55978	hypothetical prote
28	45	31.9	286	2 T21545	hypothetical prote
29	45	31.9	314	2 C97332	ABC-type transport

30	45	31.9	367	2 F96827	protein F20B17.8 (
31	45	31.9	391	2 T35698	hypothetical prote
32	45	31.9	516	2 JC6015	L- amino-acid oxid
33	45	31.9	538	2 C71946	CTP synthase - Hel
34	45	31.9	538	2 B64563	CTP synthase - H
35	45	31.9	591	2 S73708	MG31 homolog H08
36	44.5	31.6	105	2 F90939	hypothetical prote
37	44.5	31.6	105	2 B85788	hypothetical prote
38	44.5	31.6	297	1 D64724	caritine racemase
39	44.5	31.6	297	2 G85484	caritine racemase
40	44.5	31.6	297	2 G90633	caritine racemase
41	44.5	31.6	471	2 A46400	segment polarity p
42	44	31.2	152	2 A10726	probable membrane
43	44	31.2	267	2 A11249	conserved hypotet
44	44	31.2	267	2 AE1612	conserved hypotet
45	44	31.2	330	2 G95187	conserved domain p

ALIGNMENTS

RESULT 1

AG1271 hypothetical protein lmo1575 [imported] - Listeria monocytogenes (strain EGD-e)

C/Species: Listeria monocytogenes

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C/Accession: AG1271

R/Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.;

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A/Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Mat

ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,

A/Title: Comparative genomics of Listeria species

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: AG1271

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-311 <Gla>

A/Cross-references: UNIPROT:Q8Y6V6; GB:NC_003210; PIDN:CA09653.1; PID:g16411004; GSPDB:C

A/Experimental source: strain EGD-e

C/Genetics:

A/Gene: lmo1575

C/Supfamily: Mycoplasma conserved hypothetical protein MG190

Query Match 37.9%; Score 53.5; DB 2; Length 311;

Best Local Similarity 37.5%; Pred. No. 1.8;

Matches 12; Conservative 5; Mismatches 10; Indels 5; Gaps 1;

QY 1 DRVAIYEEFLRMTRN-----GTQLQNFPLDRS 27
Db 181 DRPALYRELYELPKNTVKSGLTLQNFWDEN 212

RESULT 2

A11633 hypothetical protein homolog lln1610 [imported] - Listeria innocua (strain Ctip11262)

C/Species: Listeria innocua

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C/Accession: A11633

R/Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, .;

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.;

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A/Authors: Krefl, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Mat

ok, C.; Schueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,

A/Title: Comparative genomics of Listeria species

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: A11633

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-311 <Gla>

A/Cross-references: UNIPROT:Q928B0; GB:AL592022; PIDN:CA096841.1; PID:g16414097; GSPDB:G

Best Local Similarity 45.8%; Pred. No. 26;
Matches 11; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
QY 1 DRVAIEFLRMTRNGTOLONFTL 24
DB 9 DNMRVIDFDEMDRNAHVOVKOTL 32

RESULT 7

B43332
glutamate decarboxylase (EC 4.1.1.15) beta - Escherichia coli (strain K-12)
NAlternate names: L-glutamate 1-carboxy-lyase
CSpecies: Escherichia coli
CDate: 10-Mar-1994 #sequence_revision 23-Mar-1995 #ext_change 09-Jul-2004
CAccession: B43332; S30261; H64902
RSmith, D.K.; Kassam, T.; Singh, B.; Elliott, J.F.
J. Bacteriol. 174, 5820-5826, 1992
A>Title: Escherichia coli has two homologous glutamate decarboxylase genes that map to d
A:Reference number: A43332; MUID:92394884; PMID:1522060
A:Accession: B43332
A:Molecule type: DNA
A:Residues: 1-466 <SMI>

A:Cross-references: UNIPROT:P28302; GB:M84025; NID:G146059; PIDN:AAA23834.1; PID:G146060
R.Yoshida, T.; Ueguchi, C.; Yamada, H.; Mizuno, T.
Mol. Gen. Genet. 237, 113-122, 1993
A>Title: Function of the Escherichia coli nucleoid protein, H-NS: molecular analysis of
A:Reference number: S30261; MUID:93204884; PMID:8455549
A:Accession: S30261

A:Molecule type: protein
A:Residues: 1-4, 'LOVL', 7-15 <YOS>
R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64902
A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-466 <BLAT>

A:Cross-references: GB:AE000246; GB:U00096; NID:G1787764; PIDN:AACT4566.1; PID:G1787769;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:

A:Gene: gadB
A:Map position: 78 min
C:Function:
A:Description: catalyzes the formation of gamma-aminobutyric acid, from L-glutamic acid
A>Note: In E. coli, two isoforms (alpha and beta) have been found, each encoded by a sep
C:Superfamily: Escherichia coli glutamate decarboxylase
C:Keywords: carbon-carbon lyase; carboxy-lyase; homohexamer; phosphoprotein; pyridoxal p
F/276/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 33.7%; Score 47.5; DB 1; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 3 VAIEEFLRMTRNG-TOLON 21
DB 325 IAOYTEFLRLRGEGITKVON 344

RESULT 8

S24234
glutamate decarboxylase (EC 4.1.1.15) alpha - Escherichia coli (strain K-12)
NAlternate names: GAD alpha protein; L-glutamate 1-carboxy-lyase
CSpecies: Escherichia coli
CDate: 22-Jan-1993 #sequence_revision 05-Dec-1997 #ext_change 09-Jul-2004
CAccession: S47737; H65149; S24234; S24421; A43332; PNO616
R.Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47737
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-466 <PLU>

A:Cross-references: UNIPROT:P80063; EMBL:U00039; NID:G466582; PIDN:AA18493.1; PID:G4666
R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65149

A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-466 <BLAT>

A:Cross-references: GB:AE000428; GB:U00096; NID:G1789931; PIDN:AACT6542.1; PID:G1789934;
A:Experimental source: strain K-12, substrain MG1655
R.Hars, B.; Sweeney, G.; Barra, D.; Bossa, F.; John, R.A.
Eur. J. Biochem. 204, 93-98, 1992
A>Title: The amino acid sequence of glutamate decarboxylase from Escherichia coli. Evolut
A:Reference number: S23421; MUID:92155241; PMID:1740158
A:Accession: S24234

A:Molecule type: protein
A:Residues: 1-63, 'S', 65-72, 'R', 74-152, 'N', 154-164, 'S', 166-354, 'N', 356-466 <MAR>
A:Accession: S23421

A:Molecule type: DNA
A:Residues: 148-164, 'S', 166-207, 'N', 209-294, 'V', 296-431 <MAR1>

A:Cross-references: GB:X63123; NID:G41601; PIDN:CAA44834.1; PID:G938166
R.Smith, D.K.; Kassam, T.; Singh, B.; Elliott, J.F.
J. Bacteriol. 174, 5820-5826, 1992
A>Title: Escherichia coli has two homologous glutamate decarboxylase genes that map to d
A:Reference number: A43332; MUID:92394884; PMID:1522060

A:Molecule type: DNA
A:Accession: A43332
A:Molecule type: DNA
A:Residues: 1-466 <SMI>

A:Cross-references: GB:M84024; NID:G146057; PIDN:AAA23833.1; PID:G146058
A>Note: sequence extracted from NCBI backbone (NCBIN:112979; NCBI:P112980)
R.Yoshida, T.; Yamashino, T.; Ueguchi, C.; Mizuno, T.
Biosci. Biotechnol. Biochem. 57, 1568-1569, 1993

A>Title: Expression of the Escherichia coli dimorphic glutamic acid decarboxylases is reg
A:Reference number: PNO616; MUID:94033862; PMID:7764225
A:Accession: PNO616

A:Molecule type: protein
A:Residues: 1-21, 'X' <YOS>

C:Genetics:
A:Gene: gadA
A:Map position: 78 min

C:Function:
A:Description: catalyzes the formation of gamma-aminobutyric acid, from L-glutamic acid
A>Note: In E. coli, two isoforms (alpha and beta) have been found, each encoded by a sep
C:Superfamily: Escherichia coli glutamate decarboxylase
C:Keywords: carbon-carbon lyase; carboxy-lyase; homohexamer; phosphoprotein; pyridoxal p
F/276/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 33.7%; Score 47.5; DB 1; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 3 VAIEEFLRMTRNG-TOLON 21
DB 325 IAOYTEFLRLRGEGITKVON 344

RESULT 9

E91178
glutamate decarboxylase isozyme (imported) - Escherichia coli (strain O157:H7, substrain
CSpecies: Escherichia coli
CDate: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 09-Jul-2004
CAccession: E91178
R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A89629; MUID:21156231; PMID:11258796
A:Accession: E91178
A>Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-466 <HAV>
A:Cross-references: UNIPROT:P58228, GB:BA000007, PIDN:BA837820.1, PID:G13363871, GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: EC64397
C:Superfamily: Escherichia coli glutamate decarboxylase

Query Match          33.7%; Score 47.5; DB 2; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

OY      3 VAIEEFLRMTRNG-TOLON 21
Db      325 IAQYEFRLRGEGYTKVON 344

RESULT 10
F86024
glutamate decarboxylase isozyme [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: F86024
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F86024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <STO>
A:Cross-references: UNIPROT:P58228, GB:AE005174, NID:G12518229, PIDN:AGS58658.1; GSPDB:G
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: gadA
C:Superfamily: Escherichia coli glutamate decarboxylase

Query Match          33.7%; Score 47.5; DB 2; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

OY      3 VAIEEFLRMTRNG-TOLON 21
Db      325 IAQYEFRLRGEGYTKVON 344

RESULT 11
G85726
glutamate decarboxylase isozyme [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: G85726
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85726
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <STO>
A:Cross-references: UNIPROT:P28302, GB:AB005174, NID:G12515184, PIDN:AGS56275.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: gadB
C:Superfamily: Escherichia coli glutamate decarboxylase

Query Match          33.7%; Score 47.5; DB 2; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

OY      3 VAIEEFLRMTRNG-TOLON 21

```

```

Db      325 IAQYEFRLRGEGYTKVON 344

RESULT 12
B90891
glutamate decarboxylase isozyme [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: B90891
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Sasawara, N.; Yasunaga, T.; Kunera, S.; Shiba, T.; Hatori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-466 <HAV>
A:Cross-references: UNIPROT:P28302, GB:BA000007, PIDN:BA835521.1, PID:G13361564, GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: EC62098
C:Superfamily: Escherichia coli glutamate decarboxylase

Query Match          33.7%; Score 47.5; DB 2; Length 466;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

OY      3 VAIEEFLRMTRNG-TOLON 21
Db      325 IAQYEFRLRGEGYTKVON 344

RESULT 13
T32146
hypothetical protein Cl3A2.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32146
R:Roßling, T.; Wohlmann, P.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid Cl3A2.
A:Reference number: Z21126
A:Accession: T32146
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-231 <ROH>
A:Cross-references: UNIPROT:O16875, EMBL:AF022967, PIDN:AAB69881.1; GSPDB:GN00023; CESP:
A:Experimental source: strain Bristol N2; clone Cl3A2
C:Genetics:
A:Gene: CESP:Cl3A2.11
A:Map position: 5
A:Introns: 7/3; 33/3; 71/2; 219/1

Query Match          33.3%; Score 47; DB 2; Length 231;
Best Local Similarity 47.4%; Pred. No. 13;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY      9 FLBMTRNGTOLONFLDLS 27
Db      4 FVKRTQNGKLKLPNFTVTRRS 22

RESULT 14
AB2868
UDP-hexose transferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2868
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

```

A; Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.; Oster, E.W.

A; Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A; Reference number: AB2577. MIMD.21608550. PMID.11742193

A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB2869

A;Accession: AB2868

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-316 <KUR>

A: Cross-references: UNIPROT:Q8UCW1; GB:AE008688; PIDN:AAL43360.1; PID:g177440855; GSPDB:G177440855
A: Experimental source: strain C58 (Dupont)

C;Genetics:

C;Genetics:

A; Gene: exon

A; Map position: circular chromosome

C;Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match	33.3%;	Score 47;	DB 2;	Length 316;
-------------	--------	-----------	-------	-------------

Best Local Similarity 38.5%; Pred. No. 19;

Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 DRVAIYEEFLMRNGTQLQNFLLDR 26

Db 123 DVAVYGDYERIDENGTKIGRNLIR 148

RESULT 15

E97644

UDP-hexose transferase (AF039306) [imported] - Agrobacterium tumefaciens (strain C58, ce

C;Species: *Agrobacterium tumefaciens*

C;Date: 30-Sep-2001 #sequence 30-Sep-2001 #text_change 09-Jul-2004

C;Accession: F97644

R; Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wolfram, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*
A;Reference number: A97359; MWID:21608551; PMID:11743194

A;Accession: F97644

A;Status: prelimina

A;Molecule type: DNA

A;Residues: 1-316 <KUR>

A; Cross-references: UNI

C;Genetics:

A;Gene: AGR_C_4307

A;Map position: ci

C;Superfamily: Neisseria meningitidis

Query Match	33.3%;	Score 47;	DB 2;	Length 316;
-------------	--------	-----------	-------	-------------

Best Local Similarity 38.5%; Pred. No. 19;

Matches	10;	Conservative	6;	Mismatches	10;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

QY 1 DRVAIYEEFLRMTNGTOLQNFYDR 26

Db 123 DVAVYGDYERIDENGTKIGRNLIIR 148

Search completed: October 27, 2005, 18:57:28
Job time : 18.3209 secs

123456789

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2005, 17:34:49 ; Search time 81.3953 seconds
(without alignments)
176.155 Million cell updates/sec

Title: US-10-612-090-18
Perfect score: 141
Sequence: 1 DRAIYEEFLRMTRNGTOLQNFLLDRSS 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141	100.0	1148	2 Q9H7S7	Q9H7S7 homo sapien
2	141	100.0	6995	2 Q96RK2	Q96RK2 homo sapien
3	141	100.0	22152	2 Q8WXT7	Q8WXT7 homo sapien
4	125	88.7	258	2 Q9D1H1	Q9D1H1 mus musculu
5	56	39.7	867	2 Q6ZQW5	Q6ZQW5 homo sapien
6	54	38.3	419	2 Q6UANG	Q6UANG tetradon n
7	53.5	37.9	311	2 Q8Y6V6	Q8Y6V6 listeria m
8	53.5	37.9	311	2 Q92BEO	Q92BEO listeria m
9	53.5	37.9	311	2 Q71292	Q71292 listeria m
10	53	37.6	497	2 Q726P9	Q726P9 desulfovibr
11	52	36.9	530	2 Q7Q949	Q7Q949 anopheles g
12	51.5	36.5	631	2 Q8SVW5	Q8SVW5 encephalit
13	51	36.2	186	2 Q6SDU9	Q6SDU9 bacillus l
14	51	36.2	339	2 Q9FHS7	Q9FHS7 arabidopsis
15	51	36.2	591	2 Q9LWC3	Q9LWC3 oryza sativ
16	51	36.2	1059	2 Q7RK18	Q7RK18 plasmodium
17	50	35.5	388	2 Q74GS7	Q74GS7 geobacter s
18	49.5	35.1	870	2 Q7QAH7	Q7QAH7 anopheles g
19	49	34.8	540	2 Q81390	Q81390 nicotiana t
20	49	34.8	553	2 Q8QFT1	Q8QFT1 muil cepha
21	49	34.8	766	1 PQOP_PSEPK	Q884Y3 pseudomonas
22	48.5	34.4	273	2 Q8CKM1	Q8CKM1 staphylococ
23	48	34.0	145	2 Q8PLV2	Q8PLV2 xanthomonas
24	48	34.0	191	2 Q81JQ8	Q81JQ8 plasmodium
25	48	34.0	248	2 Q9CQJ0	Q9CQJ0 m mus muscu
26	48	34.0	281	2 Q8YYP9	Q8YYP9 anabaena sp
27	48	34.0	282	2 Q76M63	Q76M63 pseudomonas
28	48	34.0	384	2 Q87DE6	Q87DE6 xyella fas
29	48	34.0	384	2 Q9PD55	Q9PD55 xyella fas
30	48	34.0	575	2 Q949V9	Q949V9 arabidopsis
31	48	34.0	575	2 Q9SKR2	Q9SKR2 arabidopsis

32	48	34.0	1109	2 Q6FJ53	Q6FJ53 candida gla
33	48	34.0	1789	2 Q7SGG0	Q7SGG0 neosporea
34	47.5	33.7	466	1 DCEA_ECO57	P58228 escherichia
35	47.5	33.7	466	1 DCEA_ECOL1	P80063 escherichia
36	47.5	33.7	466	1 DCEB_ECOL1	P28302 escherichia
37	47.5	33.7	466	2 Q7UAY1	Q7UAY1 shigella fl
38	47.5	33.7	487	2 Q83PR1	Q83PR1 shigella fl
39	47.5	33.7	489	2 Q8FHS5	Q8FHS5 escherichia
40	47.5	33.7	572	2 Q9VXL9	Q9VXL9 drosophila
41	47.5	33.7	867	2 Q6L881	Q6L881 eutrepria v
42	47	33.3	179	2 Q6Z4B4	Q6Z4B4 oryza sativ
43	47	33.3	271	2 Q99JY9	Q99JY9 mus musculu
44	47	33.3	286	2 Q93TW5	Q93TW5 stigmatella
45	47	33.3	316	2 Q8UCW1	Q8UCW1 agrobacteri

ALIGNMENTS

RESULT 1

Q9H7S7 PRELIMINARY: PRT: 1148 AA.

AC Q9H7S7, 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ14303.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta; DOI=10.1038/ng1285;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Okabayashi T., Nishitani T., Shibahara T., Tanaka T., Iishi S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tani H., Kimata M., Watanabe S., Hirooka S., Chiba Y., Ichida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fuji A., Hara R., Tanase T., Nomura Y.,
RA Togaya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima N., Sano S.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiyawa T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togsashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Iogai T., Sugano S.;
RT Complete sequencing and characterization of 21,243 full-length human
cDNAs.
RT Nat. Genet. 36:40-45(2004).
DR EMBL; AK024365; BAB14899.1; -.
DR HSSP; O9DHL1; 11VZ.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 7.
DR PROSITE; PS50024; SEA; 3.
SQ SEQUENCE 1148 AA; 127957 MW; 3861B0D5EFDFA8C CRC64;

```

Query Match          100.0%; Score 141; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 7.6e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 DRVAIEEFLLMTNNGTOLQNFLLDRSS 28
Db      1044 DRVAIEEFLLMTNNGTOLQNFLLDRSS 1071

RESULT 2
O96RK2      PRELIMINARY; PRT; 6995 AA.
AC O96RK2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Mucin 16 (Fragment).
GN Name=MUC16;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347965; PubMed=11369781; DOI=10.1074/jbc.M103554200;
RA Yin B.W., Lloyd K.O.;
RT "Molecular cloning of the ca125 ovarian cancer antigen. identification
as a new mucin, muc16."
RL J. Biol. Chem. 276:27371-27375(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX Lloyd K.O., Yin B.W.T.;
RL Submitted (SEP-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF161486; AAK74120.3; -.
DR HSSP; O9DIH1; 11VZ.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 20.
DR SMART; SMO0200; SEA; 10.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS50024; SEA; 6.
FT NON TER
SO SEQUENCE 6995 AA; 744958 MW; 80C797DBDF33A2B CRC64;

Query Match          100.0%; Score 141; DB 2; Length 6995;
Best Local Similarity 100.0%; Pred. No. 6.1e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 DRVAIEEFLLMTNNGTOLQNFLLDRSS 28
Db      6891 DRVAIEEFLLMTNNGTOLQNFLLDRSS 6918

RESULT 3
O8WXI7      PRELIMINARY; PRT; 22152 AA.
AC O8WXI7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Ovarian cancer related tumor marker CA125.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21646939; PubMed=11786729; DOI=10.1159/000050638;
RA O'Brien T.J., Beard J.B., Underwood L.J., Dennis R.A., Santin A.D.,
RA York L.;
RT "The CA 125 gene: an extracellular superstructure dominated by repeat
sequences."
RL Tumour Biol. 22:348-366(2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA O'Brien T.J., Underwood L.J., Beard J.B.;
RL Submitted (OCT-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF14442; AA16513.2; -.
DR Genew; HGNC:15582; MUC16.
DR Pfam; PF01390; SEA; 51.
DR SMART; SMO0200; SEA; 23.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_2.
DR PROSITE; PS50024; SEA; 11.
SO SEQUENCE 22152 AA; 2352668 MW; B3E7BDF19997A440 CRC64;

Query Match          100.0%; Score 141; DB 2; Length 22152;
Best Local Similarity 100.0%; Pred. No. 2.3e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 DRVAIEEFLLMTNNGTOLQNFLLDRSS 28
Db      22048 DRVAIEEFLLMTNNGTOLQNFLLDRSS 22075

RESULT 4
O9DIH1      PRELIMINARY; PRT; 258 AA.
AC O9DIH1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mus musculus 18-day embryo whole body cDNA. RIKEN full-length enriched
library, clone:111000814 product:hypothetical SEA domain containing
protein, full insert sequence.
GN Name=111000814R1k;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RL MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RL MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN PANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RA The PANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RL MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=whole body;
RL MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Kitsuurai T., Tashiro H., Itoh M.,

```



```
RA Sumi N., Ithit Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashtwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wachihi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kanai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RT Genome Res. 10:1157-1771(2000).
[6]
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=whole body;
RA Adachi J., Aizawa K., Akahira S., Akinura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Furumori Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
RA Imocani K., Ithit Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RU EMBL; AK003577; BAB22869.1; -.
DR PDB; 1IVZ; NMR; A=60-190.
DR MGD; MGI:1920982; 111008114R1k.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KM Hypothetical protein.
SQ SEQUENCE 258 AA; 29425 MW; B64D9B63394D84E7 CRC64;

Query Match 89.7%; Score 125; DB 2; Length 258;
Best Local Similarity 89.3%; Pred. No. 4,3e-11;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 154 DRAVYEEFLMTRNGTOLQNFLLDRKS 181

RESULT 5
Q6ZOW5 PRELIMINARY; PRT; 867 AA.
AC Q6ZOW5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ46845.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
(1)
RN RN
RP SEQUENCE FROM N.A.
RC TISSUE=uterus;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohara N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumoto Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sakine M.,
RA Kikuchi H., Kanda K., Nagatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugino S., Negahari K., Masuho Y., Nagai K., Isegaki T.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RU EMBL; AK126681; BAC87568.1; -.
DR InterPro; IPR000082; SEA.
DR Pfam; PF01390; SEA; 4.
DR PROSITE; PS50024; SEA; 2.
SQ SEQUENCE 867 AA; 96200 MW; C7B3033256839622 CRC64;
```

```
QY Query Match 39.7%; Score 56; DB 2; Length 867;
Best Local Similarity 42.9%; Pred. No. 12;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 DRAVYEEFLMTRNGTOLQNFLLDRSS 28
Db 380 DRAVYEEFLMTRNGTOLQNFLLDRSS 407

RESULT 6
Q6UN6 PRELIMINARY; PRT; 419 AA.
AC Q6UN6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Class I helical cytokine receptor number 20.
GN Name=CRPA20;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
NCBI_TaxID=99883;
(1)
RN RN
RP SEQUENCE FROM N.A.
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Kacinka M., Vacherie B.,
RA Blumont C., Skallil Z., Catroico L., Poulin J., Berardini Vd.,
RA Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigs R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quilley F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Crolius H.R.;
RT "Analysis of the Tetraodon nigroviridis genome reveals the
RT proteome of bony vertebrates and its duplication in teleost
RT fish."
RL Nature 0:0-0(2004).
RU EMBL; AY374492; AAR25683.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003531; Hemtreceptor_F1.
DR PROSITE; PS01355; HEMATOPOI_REC_S_F1; UNKNOWN_1.
KM Receptor.
SQ SEQUENCE 419 AA; 46687 MW; 0F100D2F20EA4FF7 CRC64;

Query Match 38.3%; Score 54; DB 2; Length 419;
Best Local Similarity 51.9%; Pred. No. 10;
Matches 14; Conservative 5; Mismatches 4; Indels 4; Gaps 2;

QY 6 YEEFLR---MTRNGT-OLQNFLLDRSS 28
Db 146 YEEFLR---MTRNGT-OLQNFLLDRSS 172

RESULT 7
Q8YV6 PRELIMINARY; PRT; 311 AA.
AC Q8YV6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE lmo1575 protein.
GN lmo1575 protein.
OS OrderedlocusNames=lmo1575;
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_TaxID=1639;
(1)
```


RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidson T.M., Zafar N., Zhou L., Radue D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Uterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.,
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
Desulfovibrio vulgaris Hildenborough";
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; AE017319; AAC97528.1; -.
DR TIGR; DVU3057; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0003824; Catalytic activity; IEA.
DR GO; GO:0004109; Ficolipoprotein oxidase activity; IEA.
DR GO; GO:0005506; Iron ion binding; IEA.
DR GO; GO:0006779; Porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; ELP3/MaB/NiB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_Tel.
DR InterPro; IPR000169; Pept_cys_acsite.
DR InterPro; IPR007197; Radical SAM.
DR InterPro; IPR011060; RlpB_bind_barrel.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; ELP3; 1.
DR TIGRFAMs; TIGR00539; hemN_tel; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 497 AA; 54126 MW; 11D323B1AFFA70DB CRC64;

Query Match 37.6%; Score 53; DB 2; Length 497;
Best Local Similarity 45.8%; Pred. No. 18;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 3 VAAYEEFLMTNGTQLONFITDR 26
Db 443 VNRDGYLRKTRNGMLVNSILR 466

RESULT 11
ID 070949 PRELIMINARY; PRT; 530 AA.
AC 070949;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCC4398 (Fragment).
GN Name=agcG50325; ORFNames=ENSANG0000010154;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RA Submitted (Mar-2002) to the EMBL/Genbank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008905; EAA09676.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR007928; Antifreeze_CF.
DR InterPro; IPR002159; CD36.
DR Pfam; PF01130; CD36; 1.
DR PRINTS; PR01609; CD36FAMILY.
FT NON_TER 1
SQ SEQUENCE 530 AA; 59511 MW; 9F8C26FC42488FF7 CRC64;

Query Match 36.9%; Score 52; DB 2; Length 530;
Best Local Similarity 69.2%; Pred. No. 28;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 12 MTRNGTQLONFITL 24
Db 257 MTRNGTQLONFITL 269

RESULT 12
ID 085W5 PRELIMINARY; PRT; 631 AA.
AC 085W5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein EC004_0440.
GN Name=EC004_0440;
OS Encephalitozoon cuniculi GB-ML.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=284813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-ML;
RX MEDLINE=21576510; PubMed=11719806; DOI=10.1038/35106579;
RA Katsinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Premier G., Barde V., Peyretallade E., Broctier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
Encephalitozoon cuniculi";
RT Nature 414:450-453(2001).
RL Nature 414:450-453(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-ML;
RA GenomeScope;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL590444; CAD25231.1; -.
KW Hypothetical protein.
SQ SEQUENCE 631 AA; 73302 MW; F8E893FABE2C5F25 CRC64;

Query Match 36.5%; Score 51.5; DB 2; Length 631;
Best Local Similarity 39.3%; Pred. No. 41;
Matches 11; Conservative 6; Mismatches 4; Indels 7; Gaps 1;

Qy 1 DRAAYEEFLMTNGTQLONFITDRSS 28
Db 375 DRAAYEEFLMTNGTQLONFITDRSS 395

RESULT 13
ID 065D9 PRELIMINARY; PRT; 186 AA.
AC 065D9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE YWKC (Putative DNA binding).
GN Name=ywkc; Synonyms=raca; ORFNames=BL04032, BL103951;
OS Bacillus licheniformis DSM 13.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=279010;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 13;
RX PubMed=1583718;
RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,
RA Ehrenreich P., Baumeister S., Henne A., Liesegang H., Mehl R.,
RA Ehrenreich A., Gottschalk G.;
RT "The Complete Genome Sequence of Bacillus licheniformis DSM13, an
RT Organism with Great Industrial Potential";
RT J. Mol. Microbiol. Biotechnol. 7:204-211(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14580;
RA Rey M.W., Ramaiya P., Nelson B.A., Brody-Karpin S.D., Zaretsky E.J.,
RA Tang M., de Leon A.L., Xiang H., Guest V., Clausen I.G., Olsen P.B.,

QY	Db	1	DRVAIVEEFLLMTRNGTQLOMFPTLDR 26	Matches 10;	Conservative 6;	Mismatches 10;	Indels 0;	Gaps 0;
Db		41	DDAILKEVKEQLKNGTIPMDQTVAKR 66					
RESULT 14								
Q9FH57			PRELIMINARY;					
AC	Q9FH57		PRT;					339 AA.
DT	01-MAR-2001		(TREMBLrel. 16, Created)					
DT	01-MAR-2001		(TREMBLrel. 16, Last sequence update)					
DT	05-JUL-2004		(TREMBLrel. 27, Last annotation update)					
DE	GATA-binding transcription factor-like protein (A5566320).							
GN	Name=A5566320;							
OS	Arabidopsis thaliana (Mouse-ear cress).							
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae;							
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsi.							
OX	NCBI_Taxid=3702;							
	[1]							
RP	SEQUENCE FROM N.A.							
RX	MEDLINE=20181125; PubMed=10718197;							
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,							
RA	Tabata S.;							
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence							
RT	features of the regions of 3,076,755 bp covered by sixty P1 and TAC							
RL	DNA Res. 7:31-63(2000).							
	[2]							
RP	SEQUENCE FROM N.A.							
RA	Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,							
RA	Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,							
RA	Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,							
RA	Kawai J., Kim C.J., Narusaka M., Quach H.L., Sakurai T., Satou M.,							
RA	Seki M., Shim P., Tang C.C., Torouni M., Wallender B.K., Wong C.,							
RA	Wu H.C., Yamada K., Yu G., Yuan S., Shinozaki K., Becker J.,							
RA	Theologis A., Davis R.W.;							
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.							
RP	[3]							
RP	SEQUENCE FROM N.A.							
RA	Kim C.J., Chen H., Cheuk R., Shim P., Bowser L., Carninci P.,							
RA	Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,							
RA	Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,							
RA	Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,							
RA	Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,							
RA	Shinozaki K., Davis R.W., Theologis A., Becker J.R.;							
RL	Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.							
DR	EMBL	AB022211	BAB10711.1	-				
DR	EMBL	AY136450	AA097115.1	-				
DR	EMBL	BT010367	AA					

```

Qy      5  IYEEFLRMT--RNGTOLQNFLLD 25
      :|||||:|||||:|||||:
Db      22  YVEEFLAVTTAQNQNGFSVDDFSD 44

RESULT 15
Q9LWC3      PRELIMINARY;      PRT;      591 AA.
AC      Q9LWC3
DT      01-OCT-2000 (TrEMBLrel. 15, Created)
DT      01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE      01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE      ESTs C19814 (E109711).
OS      Oryza sativa (japonica cultivar-group).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCBI_Taxid=39947;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      PubMed=12447438; DOI=10.1038/nature01184;
RA      Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA      Wu J., Milmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA      Hoshikawa S., Masubuku M., Arikawa K., Chiden Y., Hayaishi M.,
RA      Okamoto M., Ando T., Aoki H., Arita K., Hanada M., Harada C.,
RA      Hishinaka S., Honda M., Ichikawa Y., Iiduma A., Iijima M., Ikeda M.,
RA      Ikemoto M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA      Katsawa M., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA      Machida K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA      Nagasaki H., Nakashima M., Nakaya Y., Nakamichi Y., Nakamura M.,
RA      Namiki N., Negishi M., Ohta I., Ono N., Sai S., Sakai K., Shibata M.,
RA      Shimokawa T., Shomura A., Song J., Takaraki Y., Terasawa K., Tsuji K.,
RA      Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshinara R., Yukawa K.,
RA      Zhang H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA      Zhao M., Jiang J., Gotohorai T.;
RL      *The genome sequence and structure of rice chromosome 1." ;
RT      Nature 420:312-316(2002).
DR      EMBL; AP002094; BAA96221.1; -
DR      Gramene; Q9LWC3; -
DR      InterPro; IPR004993; GH3.
DR      Pfam; PF03321; GH3; 1.
SQ      SEQUENCE      591 AA; 64190 MW; 9A4D092E7B28D39B CRC64;

Query Match      36.2%; Score 51; DB 2; Length 339;
Best Local Similarity 47.8%; Pred. No. 24;
Matches 11; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

4  AIYEEFLRMTNRNGTOLQNFLLD 26
:|||||:|||||:|||||
19  AVIEEFERVRTRDANNVQRETLRR 41

```

Search completed: October 27, 2005, 18:55:24
Job time : 84.3953 sec